GEOMAR
Helmholtz Centre for Ocean Research Kiel

future ocean
KIEL MARINE SCIENCES

MSc Biological Oceanography
Master Thesis

Invasion Genomics: Population structure and diversity patterns in the invasive ctenophore Mnemiopsis leidyi based on whole-genome re-sequencing.

Moritz Ehrlich

June 2017

1st Supervisor: Dr. Cornelia Jaspers
2nd Supervisor: Prof. Thorsten Reusch

Abstract

Increased global connectivity has led to an elevated prevalence of invasive species. Marine invaders in particular have caused significant damage to both ecosystems and economies outside of their native ranges. The mechanisms by which ecological, genetic and demographic factors combine to confer invasion success have yet to be understood. Specifically the patterns and importance of genetic diversity during colonisation, establishment and expansion have been debated in recent years. Here we describe population structure and contrast genomic diversity between native and invasive populations of a marine invader for the first time at the whole-genome scale. We examine the Mnemiopsis leidyi invasion system, sampling two native sites along the Western Atlantic and three invasive sites in the North, Black and Western Mediterranean Seas. Whole-genome re-sequencing of 72 specimens produced a mean coverage of 26x per individual and resulted in over 6.5 million SNPs. We confirm two independent invasion events from diverged source regions: the first from the vicinity of the Floridian Peninsula into the Black Sea with a subsequent range expansion throughout the Mediterranean, and the second from the coast of New England into the North Sea. Inbreeding coefficients and tests of HWE suggest M. leidyi populations are in equilibrium and do not self-fertilise in the native or the invasive range. Comparisons of invasive and source populations display patterns of increase, maintenance and decrease of genetic diversity at the whole-genome scale. We reject the former paradigm of diversity reduction through founder effects in invasive populations and hypothesise that propagule pressure or unrelated demographic processes both in invasive and native ranges predominantly shape genomic diversity. Claims are supported by demographic indicators and estimates of effective population size derived from the genome-wide SNP set.

Contents

1. Introduction

1.1 Background

Spurred by anthropogenic interference, ecosystems around the globe have become increasingly connected. The multitude of highly frequented international trading routes has allowed species to expand their habitat range beyond the potential of their inherent dispersal capacities and establish non-indigenous populations far from their native sites (Carlton and Geller, 1993; Wonham *et al.*, 2001). Whilst some of these non-indigenous species (NIS) integrate into existing ecosystems and may even provide certain benefits (Gozlan, 2008; Schlaepfer, Sax and Olden, 2011), many colonising organisms oftentimes cause severe ecological or economic damage, at which point they are termed invasive species (Pimentel, 2011; Simberloff *et al.*, 2013). As a general rule-of-thumb it can be assumed that one in ten species arriving to a new habitat will establish, whereas a tenth of those established will become invasive (Williamson and Fitter, 1996).

Marine biological invasions have attracted growing attention owing to their extensive threat to ecosystems worldwide (Crooks, 2002; Bax *et al.*, 2003; Chown *et al.*, 2015), the global economy (Pejchar and Mooney, 2009; Pimentel, 2011) and in certain cases even human health (Tatem, Hay and Rogers, 2006). With the steady rise of global shipping, marine ecosystem connectivity and hence the prevalence of marine NIS has drastically increased (Hulme *et al.*, 2008). Undeniably, a rigorous analysis of the invasion vectors, dynamics and susceptibility parameters in the marine environment will be crucial in mitigating current and preventing future scenarios (Hewitt and Campbell, 2007; Seebens, Gastner and Blasius, 2013). A fundamental understanding of marine invasion dynamics may also give us valuable insights into the biological responses to climate change in aquatic ecosystems (Sorte *et al.*, 2013; Moran and Alexander, 2014). Anthropogenic habitat fragmentation and rapidly shifting environmental conditions could make the colonisation of distant yet favourable habitats a more advantageous strategy than costly local adaptation, possibly leading to a higher prevalence of marine invaders in the near future (Colautti and Lau, 2015).

Contemporary environmental niche models designed to predict invasive ranges based on tolerance limits in the native habitat are only successful to a degree (Blakeslee *et al.*, 2010; Tepolt, 2015). It has become apparent that rapid local adaptation plays a major role in successful biological invasions, imploring the development of evolutionary as well as ecological models (Whitney and Gering, 2015). A paradigm shift, characterising not species but populations based on their genetic composition, adaptive potential and connectivity to novel habitats will be necessary to reliably manage invasion hazards in the future (Tepolt, 2015).

1.2 Invasion Genetics and Diversity

Low genetic diversity in the founding population was thought to be a critical impediment to invasive success (Baker and Stebbins, 1965; Colautti and Lau, 2015; Rius *et al.*, 2015). Finite quantities of colonising individuals restrict the number of genotypes whilst strong drift further erodes diversity over time in small founder populations. Inbreeding depression and the lack of adaptation potential due to a restricted gene pool should eradicate early colonies relatively quickly (Allendorf and Lundquist, 2003; Dlugosch and Parker, 2008; Gaither, Toonen and Bowen, 2012; Riquet *et al.*, 2016). Nonetheless, there exist several examples of genetically deprived yet highly successful invaders (Darling *et al.*, 2008; Richards, Schrey and Pigliucci, 2012). This so called genetic paradox of invasions (Sax and Brown, 2000) is hypothesised to result from strong drift during extended lag phases (Gaither, Toonen and Bowen, 2012; Bock *et al.*, 2015), close genotype matching to local environmental parameters (Verhoeven *et al.*, 2011), extreme phenotypic plasticity (Richards, Schrey and Pigliucci, 2012), incompatibilities between differentiated source populations (Dlugosch *et al.*, 2015) or gene surfing (Burton and Travis, 2008; Excoffier, Foll and Petit, 2009).

The most recent discourse in the field of invasion genetics acknowledges the existence of low-diversity NIS as well as the underlying processes leading to their success but argues that they are the exception (Roman and Darling, 2007; Estoup *et al.*, 2016). After all, genetic variety in invasive populations brings with it significant short-term (e.g. hybrid vigour) (Drake, 2006; Arnold, Ballerini and Brothers, 2012) and long-term benefits (e.g. additional selection targets) (Rius and Darling, 2014; Colautti and Lau, 2015; Whitney and Gering, 2015) that should result in a higher probability of establishment. Indeed, the majority of documented invasive species display elevated genetic diversity (Roman and Darling, 2007; Rius *et al.*, 2014; Tepolt, 2015), equal to or above their native range counterparts (Kolbe *et al.*, 2004; Gillis *et al.*, 2009). This has been attributed to large inoculates, repeated introductions from a single source region (Rius *et al.*, 2015; Viard, David and Darling, 2016), multiple introductions from distinct source regions (Darling *et al.*, 2008) or inbreeding relief (Verhoeven *et al.*, 2011; Rius and Darling, 2014).

A recurring diversity pattern, driven by taxonomic membership or reproductive strategy, does not seem to exist (Rius *et al.*, 2014). This begs the question whether genetic diversity is an active contributor in conferring invasive success or outright irrelevant and just a correlate of more significant ecological or demographic factors such as propagule pressure (Rius and Darling, 2014; Szűcs *et al.*, 2014; Bock *et al.*, 2015; Dlugosch *et al.*, 2015; Whitney and Gering, 2015)(but see Forsman 2014). The density dependent increase in an individual's fitness, known as the Allee effect (Allee, 1931), seems to be pervasive in colonising populations (Taylor and Hastings, 2005) with establishment success often correlated to propagule pressure (Lockwood, Cassey and Blackburn, 2005). It is also evident that genetic diversity at a given site is stochastically linked to population size, leading to the confounding of both genetic and ecological density dependent effects (Simberloff, 2009). Although a number of studies have aimed to resolve these components experimentally (Hufbauer *et al.*, 2013; Szűcs *et al.*, 2014), disentangling genetic diversity from the underlying ecology is inherently problematic and their relative contributions towards the establishment of invasive populations remains elusive (Rius and Darling, 2014).

1.3 The *M. leidyi* Invasion System

The lobate ctenophore *Mnemiopsis leidyi* (A. Agassiz, 1865) is one of the most successful marine invaders (Lowe *et al.*, 2000) and a primary contributor to the anthropogenic "jellification" of the oceans (Purcell, Uye and Lo, 2007; Richardson *et al.*, 2009). Its introduction had significant impact on both ecosystems and the economy in its invasive range (Oguz, Fach and Salihoglu, 2008), possibly leading to the collapse of entire fisheries (Kideys *et al.*, 2005; Roohi *et al.*, 2010) (but see: Bilio and Niermann, 2004). Native to the western Atlantic (GESAMP 1997) it has invaded the Black Sea during the 1980s (Vinogradov *et al.*, 1989) and in later years the Caspian (Ivanov *et al.*, 2000), Mediterranean (Kideys and Niermann, 1993; Fuentes *et al.*, 2009), North (Faasse and Bayha, 2006) and Baltic Seas (Javidpour, Sommer and Shiganova, 2006). Like other planktonic species (Carlton and Geller, 1993; Wonham *et al.*, 2001; Briski *et al.*, 2013), the ctenophore most likely invaded via ballast water transfer of major shipping lines, the most common transport vector of marine NIS (Molnar *et al.*, 2008). Its high growth and reproduction rates (over 10,000 eggs per day)(Reeve, Syms and Kremer, 1989; Jaspers, Møller and Kiørboe, 2015), dense distribution (ca. 1 ind. L^{-1})(van Walraven, Langenberg and Van der Veer, 2013), broad diet (Costello *et al.*, 2012) and ability to self-fertilise (Jaspers, Møller and Kiørboe, 2015) make it particularly suited to colonise novel habitats from just a small number of founding individuals. Ultimately, a holoplanktonic lifecycle and broad environmental tolerance (GESAMP 1997; Purcell et al. 2001) probably allowed for efficient dispersal and rapid establishment in the invasive range as evidenced in other invertebrate taxa (Rohfritsch *et al.*, 2013).

However, in contrast to other holoplanktonic species (Stopar *et al.*, 2010; Bortolotto *et al.*, 2011) *M. leidyi* exhibits notable population structure. By using a limited number of mitochondrial and nuclear markers, earlier studies have begun to describe population structure in the native and invasive range, including the identification of putative source regions, sites of primary establishment and possible secondary introductions. Reusch *et al.* (2010) and Ghabooli *et al.* (2011) used microsatellite markers and ITS sequences

respectively to identify two well-differentiated native populations that seem to have independently invaded European waters. A southern haplogroup, native to the East Coast of the United States south of Cape Hatteras, NC and extending into the Gulf of Mexico seems to have invaded the Black and Caspian Seas, whilst a northern haplogroup occurring between Cape Hatteras, NC and New England, appears to have invaded the North and Baltic Seas. Bolte *et al.* (2013) confirmed this structure and demonstrated that the Mediterranean population is most likely the result of a range expansion from the Black Sea, though repeated introductions from the Gulf of Mexico might have occurred in parallel (Ghabooli *et al.*, 2013). Recently Bayha *et al.* (2015), using the same microsatellite markers as well as mitochondrial cytochrome b sequences, detected fine structure in the native region further differentiating the southern haplogroup into Atlantic and Gulf of Mexico sub-populations. Moreover, they identified the oceanographic front produced by the collision of the Gulf Stream and Labrador Current at Cape Hatteras, NC to be a physical barrier to admixture between the northern and southern haplogroups. Populations were found to be large with high internal gene-flow across vast spatial scales yet with recognisable signals of isolation-by-distance.

As evidenced above, *M. leidyi* displays interesting phylogeography that lends itself to the study of marine invasion genomics. The two genetically and geographically independent invasion waves can be seen as natural replicates and harnessed to investigate molecular parallelism (Jones *et al.*, 2012; Bradic *et al.*, 2013), a process not uncommon in closely related lineages (Bollback and Huelsenbeck, 2009; Conte *et al.*, 2012) experiencing similar selection pressures (Bailey, Rodrigue and Kassen, 2015). Analogous patterns of genomic diversity in both invasion waves could highlight the influence of the invasion process itself, rather than local environmental conditions or demography, in shaping genetic diversity. Since the two invasion waves are separated temporally by roughly 20 years the effect of time on the genomic diversity of colonising species can also be addressed. In fact, a documented expansion lag time (Pereladov, 1988; Vinogradov *et al.*, 1989) coupled with a hermaphroditic reproduction mode appears to have significantly reduced genetic diversity in the Black Sea invasive

population as opposed to the relatively young northern invasion (Bayha *et al.*, 2015). Furthermore, the range expansion into the Mediterranean Sea offers an opportunity to study secondary "stepping-stone" invasions and its conceivable impact on genetic diversity under the assumption that every step is comparable to a new founder event. Finally, well-characterised structure in the source regions reduces the chance of ascertainment bias whilst large, planktonic populations advocate local panmixia (Stopar *et al.*, 2010; Bortolotto *et al.*, 2011), reducing the probability of omitting fine geographic structure due to incomplete sampling (Colautti and Lau, 2015; Viard, David and Darling, 2016).

Although a select number of neutral markers was suitable for determining the underlying population structure of *M. leidyi*, genome-wide patterns could not be addressed. Accurate quantification of demographic indicators, diversity and their variability across the genome require a much greater marker resolution (Funk *et al.*, 2012; Hohenlohe *et al.*, 2013), especially in a holoplanktonic marine species where large effective population sizes and high levels of gene-flow readily erode linkage blocks (De Wit and Palumbi, 2013; Benestan *et al.*, 2015; Tepolt, 2015). Fortunately, the recent publication of the *M. leidyi* genome (Ryan *et al.*, 2013) made the development of high-density, genome-wide single nucleotide polymorphism (SNP) markers possible. With its relatively small genome (156 Mb), high proportion of genic regions (58%) and low quantity of repetitive sequences (Ryan *et al.*, 2013), *M. leidyi* is an enticing target, suited to develop the field of invasion genomics for which few model organisms exist (Chown *et al.*, 2015).

Here we employ whole-genome re-sequencing of *M. leidyi* individuals sampled from both native and invasive ranges to develop a dense set of SNP markers used to address patterns of genetic diversity and demography as a result of the invasion process. A couple of previous studies have tackled similar themes in other marine invaders, including *Crepidula fornicata* (Riquet *et al.*, 2013)*, Crassostrea gigas* (Rohfritsch *et al.*, 2013)*, Carcinus maenas* (Tepolt and Palumbi, 2015), *Gambusia holbrooki* (Díez-Del-Molino, 2016; Vera, Díez-Del-Molino and García-Marín, 2016) and *Fistularia commersonii* (Bernardi *et al.*, 2016), but have employed few, highly polymorphic markers (i.e. microsatellite markers) or reduced representation sequencing (i.e. RAD-Seq). To our knowledge this is the first study to compare native with invasive populations from two spatially and temporally independent marine invasions at the whole-genome scale.

2. Methods

2.1 Sample Collection

Sampling sites were chosen based on the previously identified broader population structure of *Mnemiopsis leidyi* in the North Atlantic, Mediterranean and Black Sea (Reusch *et al.*, 2010; Ghabooli *et al.*, 2011, 2013; Bolte *et al.*, 2013; Bayha *et al.*, 2015) (Figure 1). 74 specimens were sampled from 5 distinct locations: Woods Hole, MA and Miami, FL on the East coast of the United States, representative of the northern and southern native populations, Sylt, DE representing the recent invasion into the North Sea, Varna, BG representing the invasive population in the Black Sea and finally Villefranche-sur-Mer, FR, chosen to represent the limit of the *M. leidyi* invasive range in the western Mediterranean Sea (Supplement A). This sampling strategy covered the major North Atlantic population structure at sufficient sample size (Willing, Dreyer and van Oosterhout, 2012) and allowed for parallel contrasts within and between native and invasive ranges.

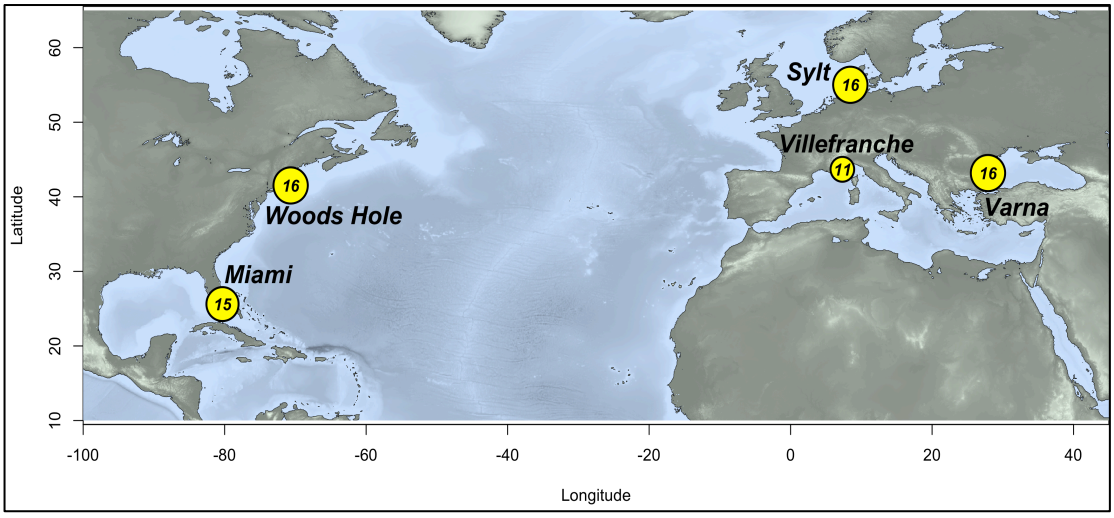

Figure 1: Map of sampling locations. Number of specimens sampled shown in insets.

Specimens were collected dockside or from small vessels which operated less than 500m from shore. All samples were taken from surface waters (<1m depth) and either kept alive or frozen below -40 °C until further processing. Entire specimens were briefly rinsed in distilled water and subsequently preserved for transportation and storage through lyophilisation (Supplement B).

2.2 DNA Isolation, Library Preparation and Sequencing

High molecular weight (HMW) DNA was isolated from the lyophilized ctenophore tissue using a modified CTAB/Chloroform protocol (Supplement C). Prior to library preparation DNA isolates were evaluated for purity and integrity using the NanoDrop spectrophotometer (Thermo-Fisher Scientific Inc., Waltham, MA) and via 0.8% agarose gel electrophoresis respectively. Paired-end libraries with an insert size of 550 bp were constructed for all 74 samples and uniquely indexed using the Illumina TruSeq Nano DNA HT Library Preparation Kit (Illumina, San Diego, CA). Libraries were subsequently quantitated and quality controlled using an Agilent 2100 Bioanalyzer System (Agilent Technologies, Santa Clara, CA). Contamination in the form of adapter-dimers was removed through gel size selection. Equimolar quantities of the processed libraries were multiplexed into two pools; each consisting of 37 uniquely indexed libraries corresponding to single *M. leidyi* individuals. Each multiplex pool was sequenced on three replicate flowcells in 2x150 bp paired-end, high-output mode on the Illumina NextSeq 500 System (Illumina, San Diego, CA), giving a theoretical total yield of 600-720 Gb equivalent to a sequencing depth of 60X per individual.

2.3 Raw Data Processing

Raw sequence data was demultiplexed separately for four of the six flowcells and converted to FASTQ format using *bcl2fastq* v.2.17.1.14 (Illumina, San Diego, CA), allowing for one mismatch in the barcode sequence. In the remaining two flowcells a sequencing error led to only 6 of the 8-base barcode being read. Although barcodes remained distinguishable, the number of allowed mismatches *i.e.* sequencing errors, was set to zero for certain barcode combinations. Prior to further processing read quality metrics, including base quality Phred scores as well as levels of read duplication and adapter contamination, were collected using *FastQC* v.0.11.5 (Babraham Bioinformatics, Cambridge, UK).

Raw sequence data was processed according to the *GATK* Best Practices pipeline for variant discovery in germline cells (Van der Auwera *et al.*, 2013) using *Picard Tools* v.2.6 (Broad Institute, Cambridge, MA) and the *Genome Analysis Toolkit* v.3.6 (McKenna *et al.*, 2010). In brief, contaminant adapter sequences within the raw reads were effectively removed by reducing base quality scores to 2. Next, reads were mapped to a trimmed *M. leidyi* reference genome using the *mem* algorithm in *BWA* v.0.7.15 (Li and Durbin, 2009). The reference genome had previously been sourced from the *EMBL-EBI* genome database and trimmed to only contain scaffolds longer than 10 Kb. This reduced the number of scaffolds from 5100 to 1254 whilst retaining 94.9% of the original sequence (Supplement D). Mapped reads were then curated using the *MergeBamAlignments* tool in *Picard*, which recovers original base quality scores, restores hard-clipped reads and adds meta information regarding read-group membership for later use. Next, sequencing duplicates were removed after which read groups belonging to the same sample were merged across all flowcells. A second round of duplicate removal was performed to exclude PCR duplicates generated during library preparation. *QualiMap* v.2.2.1 (Okonechnikov, Conesa and García-Alcalde, 2015) was used to gather mapping statistics (*e.g.* sequencing depth, mapping quality *etc.*) and identify anomalous samples. Ambiguously mapping reads as well as those with mapping quality below 20 were discarded.

2.4 SNP Calling and Filtering

Variant calling was performed by running the *GATK HaplotypeCaller* algorithm in GVCF-mode followed by joint genotyping of all samples. Collective genotyping has proven to be powerful at lower sequencing depths by harnessing multi-sample allele frequencies to identify true variants (Van der Auwera *et al.*, 2013). From the output file only nuclear SNPs with quality scores above 30 were retained; indel sites were removed and not analysed any further in this study whilst SNPs in the mitochondrial genome were discarded due to inherent differences in the analysis of polyploid data. The remaining SNPs were subsequently hard-filtered based on several quality indicators. These statistics

included i) Quality-By-Depth, a score used to normalize quality scores inflated due to deep coverage, ii) Fisher Strand Bias, an exact test of whether variant alleles are distributed evenly on forward and reverse strands, iii) Strand Odds Ratio, another strand bias statistic derived from the odds ratio test, iv) Root-Mean-Square Mapping Quality, used to assess the mapping quality of reads at the SNP site, v) Mapping Quality Rank Sum, a rank sum test statistic assessing parity of mapping qualities between reference and variant alleles and finally vi) Read Position Rank Sum, a rank sum test quantifying read position bias of variant alleles versus reference alleles. Threshold values for hard-filtering were defined through visual inspection of the empirical distribution of the respective quality metrics as suggested by the *GATK* Best Practices workflow. The distributions and resulting threshold values can be found in Supplement E.

Next, filtering was performed at the genotype level by discarding all genotypes with Phred quality scores below 13, any SNPs that became monomorphic as a result were discarded (Supplement F). Finally, the SNP set was filtered for call rate by retaining only those sites where more than 70% of individuals had been genotyped in order to reduce the amount of missing data. For analyses requiring minimal amounts of missing data as well as highly informative SNPs, an additional minor allele frequency (MAF) filter and a stricter call rate filter were implemented. These removed sites with MAF below 0.05 and with more than one missing genotype respectively. Finally, whenever analyses were conducted at the population level the above filters were applied at the population level also (Supplement G). This led to the construction of several distinct SNP sets, both lenient (70% call rate) and strict (max. 1 missing genotype & MAF < 0.05), at the global and population level.

2.5 Population Structure

A pairwise distance matrix was calculated from the strict, global SNP set containing only biallelic SNPs using the software package *poppr* v.2.3.0 (Kamvar and Grunwald, 2014) in *R* v.3.3.2 (R Core Team, 2016). A neighbour-joining tree was constructed from the distance matrix using the *bionjs* algorithm from the software package *ape* v.4.1 (Paradis, Claude and Strimmer, 2004). This algorithm is based on the *BIONJ* algorithm (Gascuel, 1996) and is optimized for missing data. Bootstrap support was calculated using the *aboot* function in *poppr* with 1000 iterations.

Principal component analyses (PCA) were conducted with the strict, global SNP set of biallelic SNPs using the package *pcadapt* v.3.0.4 (Luu, Bazin and Blum, 2017) in *R*. The number of principal components, *K*, was defined by running a preliminary PCA with $K = 15$ and applying Cattell's rule (Cattell, 1966) to the scree plot of eigenvalues. Sample projections onto predetermined principal components were achieved by normalising the genotype matrix according to Patterson, Price and Reich (2006), replacing missing genotypes by the median genotype at that locus, and multiplying its transpose by the loading vectors calculated in *pcadapt*.

After confirming population structure, all SNPs in the lenient set were tested for Hardy-Weinberg equilibrium both globally and at the population level using the exact test (Wigginton, Cutler and Abecasis, 2005) implemented in the software package *VCFtools* v.0.1.15 (Danecek *et al.*, 2011). Next, *q*-values were calculated from the resulting *p*-value distribution using the *qvalue* package v.2.8.0 (Bass, Dabney and Robinson, 2015) in *R*. SNPs were deemed to be out of Hardy-Weinberg equilibrium when testing significantly at a false-discovery rate of $\alpha = 0.05$. Significant SNPs were further sorted according to their directionality *i.e.* whether heterozygote deficit or excess was the cause of disequilibrium.

Weighted and unweighted global F_{ST} estimates (Weir and Cockerham, 1984) were calculated in overlapping windows of 10 Kb with a step size of 1 Kb using the lenient SNP set in *VCFtools*. The resulting distributions of F_{ST} estimates were plotted and their means taken as the singular, genome-wide estimates. Pairwise F_{ST} estimates between populations were calculated in a similar fashion by taking the mean of the pairwise F_{ST} estimates across all windows. One-sample t-tests were conducted on the F_{ST} distributions to obtain confidence intervals and confirm means were non-zero.

2.6 Genomic Diversity

Individual multilocus heterozygosity (IMLH) was calculated with *VCFtools* for every individual using the population-specific strict SNP sets. However, IMLH is upward biased for SNP sets derived from small sample sizes due to an underrepresentation of rare polymorphisms. To counteract said ascertainment bias, populations were randomly subsampled to their largest common sample size. Any SNPs that became monomorphic due to subsampling were discarded thus simulating their non-inclusion at low sample sizes. In addition, the SNP set was thinned to only contain loci more than 10 Kb apart in order to avoid linkage bias. Population-level heterozygosity, H_S, was estimated as the proportion of heterozygotes at a locus, averaged across all SNPs in the thinned, subsampled, population-specific set. A Welch-corrected ANOVA was run across samples of individuals to test for significant differences in IMLH between sampling sites. Games-Howell post-hoc tests were applied to identify significantly different population pairs. To assess differences in population-level heterozygosity a Kruskal-Wallis test was run across loci followed by Nemenyi post-hoc tests with Chi-squared approximation using the *PMCMR* package (Pohlert, 2014) in *R*.

Watterson's estimator, θ_W (Watterson, 1975), Nei and Li's nucleotide diversity, π (Nei and Li, 1979), and Tajima's *D* (Tajima, 1989), were computed for each population in non-overlapping windows of 10 Kb using the respective population-specific lenient SNP sets. The former statistic was computed from SNP density data in *R* whereas the latter two were derived using designated functions in *VCFtools*. Genome-wide distributions of θ_W, π and *D* were constructed and subsequently compared between populations using Kruskal-Wallis omnibus tests and Nemenyi post-hoc tests with Chi-squared approximations from the *PMCMR* package. Distributions of Tajima's *D* were compared using a Welch-corrected ANOVA followed by Games-Howell post-hoc tests. Population-specific deviations from zero were evaluated via one-sample *t*-tests.

Individual inbreeding coefficients, *F*, were calculated using Visscher's method of moments estimator described in Yang *et al.* (2010) and implemented in *VCFtools*. As with the IMLH, the thinned, population-specific, strict SNP sets were used, however, subsampling was not required as the estimator is robust to unequal sample sizes. Population-level inbreeding coefficients, F_{IS}, were estimated using a sample-size-corrected formula (Weir and Cockerham, 1984) and taking the weighted average over all loci as suggested by Reynolds, Weir and Cockerham (1983). A Welch-corrected ANOVA was run across samples of individuals to test for significant differences in *F*. Games-Howell post-hoc tests were applied to identify significantly different population pairs whilst one-sample *t*-tests were used to assess significant difference from zero.

2.7 Demographic Indicators

The global, folded minor allele frequency (MAF) distribution was retrieved using *VCFtools* with biallelic loci from the global, lenient SNP set. Population-specific MAF distributions were harder to attain since allele frequencies become discretised at low sample sizes and/or low call rates. To overcome this problem, populations were subsampled at random to their largest common sample size whilst utilising only loci with 100% call rate. By ensuring a common baseline the population-specific minor allele counts could be plotted and compared.

Population-specific linkage disequilibrium (LD) decay was analysed by computing the squared correlation coefficient, r^2, between pairs of unphased genotypes using *VCFtools*. Only biallelic loci from the respective strict, population-specific SNP sets were used with physical distances between genotype pairs varying from 500 bp to 400 Kb. Next, pairwise distances were binned into discrete blocks every 1 kb and the weighted mean of r^2-values calculated for each. Finally, the mean r^2-values were corrected for sample size as suggested by Corbin *et al.* (2012) albeit using the harmonic mean of sample sizes across all loci rather than the number of specimens. This is owing to the fact that effective sample size is lower at several loci due to incomplete genotype calling. LD decay was graphed by plotting adjusted r^2-values against physical distance in kilobases.

Effective population sizes, N_e, through time were estimated by applying the method of variable N_e developed by Corbin *et al.* (2012), using the adjusted r^2-values calculated previously. Recombination rate was assumed to be comparable to that of the invertebrate *Caenorhabditis elegans* with an average of approximately 3 cM/Mb (Prachumwat, DeVincentis and Palopoli, 2004). Recombination frequency was estimated from mapping distance using Haldane's mapping function (Haldane, 1919) and corrected according to Sved and Feldman (1973). The mutation rate correction term (Ota and Kimura, 1971) was set to $\alpha = 2$ as suggested by Tenesa *et al.* (2007). The number of generations in the past were derived from the recombination frequency according to Hayes *et al.* (2003).

3. Results

3.1 SNP Calling

3.1.1 Sequencing and Mapping

Raw sequencing output after demultiplexing came to 380 Gb with a mean base quality of 28.4 and mean GC-content of 39.7%. Total output fell short of the expected 600 Gb mainly due to residual adapter dimer impurities in 4 of the 6 flowcells. This problem was mitigated via gel size selection of libraries in flowcells 5 and 6, resulting in a 4-fold yield (Supplement H). Sequencing output was mostly uniform across populations except for a disparity in flowcell 5. Here, uneven multiplexing led to a 2-fold increase in the yield of all Sylt samples and 8 Varna samples as well as a consequential 2-fold decrease in the yield of all Villefranche and 7 Varna samples. Moreover, two Villefranche samples were excluded from further analysis due to extremely low sequencing output (<5 Mb) in one, and anomalous GC-content (70%) in the other case.

On average 87.9% of reads were successfully mapped with a mean mapping quality of 42.8, insert size of 418 bp and duplication rate of 11.7%. Given that the reference assembly had been reduced to 94.9% of its original length this translates to an effective mapping rate of 92.6% and higher mean quality scores due to some reads being inhibited from mapping. Between-population differences are likely due to the fact that the reference sequence is that of a Woods Hole specimen. Diminished mapping rates and qualities for genetically more distant individuals can therefore be expected. As such, Sylt and Woods Hole specimens showed significantly higher mean mapping qualities and rates than the other populations, although absolute differences were negligible (Supplement I).

Mean sequencing depth per individual was 26.3x with an average of 6.3% unmapped regions. Depth was uniform across major parts of the genome although smaller scaffolds displayed mean sequencing depths of over 100x; most likely artefacts caused by under-assembled duplicate regions in the reference genome (Supplement J). Ambiguously mapping reads and subsequent lower mapping qualities in small scaffolds seem to justify the purposeful size reduction of the reference genome. Differences in sequencing depth between populations arose primarily as a result of the uneven output of flowcell 5, with only minor differences due to variable mapping rates. Sylt individuals were over-sequenced at a mean depth of 36.8x whilst Villefranche specimens were under-sequenced at 14.0x. Miami, Woods Hole and Varna specimens were covered at depths close to the global mean at 24.6x, 26.8x and 23.2x respectively.

3.1.2 SNP Filtering and Subsetting

The *GATK* variant calling pipeline and subsequent hard-filtering for genotype quality resulted in a global raw set containing 7.3 million SNPs, translating to a mean density of 50.3 SNPs per Kb (Supplement K). Further filtering for call rate and MAF resulted in lenient and strict global sets described in Table 1. Lower sequencing depth in Villefranche and 7 Varna samples and the resultant diminished call rates are primarily responsible for the large discrepancy in SNP numbers between the two sets (Supplements G & J).

Table 1: Filters and SNP counts of the two global SNP sets.

Global SNP Set	***Call Rate Filter***	***MAF Filter***	***SNP Count***
Lenient	*70%*	*-*	*~ 6.75 million*
Strict	*98%*	*0.05*	*57,080*

Population-specific sets resulted in varying SNP counts, mainly due to distinct per population call rates (Figure 2). As such, the Villefranche sets contain significantly lower SNP counts as a direct result of lower sequencing depth. Whilst a small sample size (n=9) and lower power to detect private SNPs could also contribute to lower counts, this is not the case. Artificial subsampling to the largest common sample size resulted in lower SNP counts but a similar qualitative pattern (Supplement L), suggesting inferior sequencing depth, not sampling bias, to be the cause. It should be noted that population-specific SNP sets might contain monomorphic sites if a specific locus is polymorphic within another population or between populations. Figure 2 therefore does not describe the number of segregating sites per population but rather the number of successfully genotyped loci.

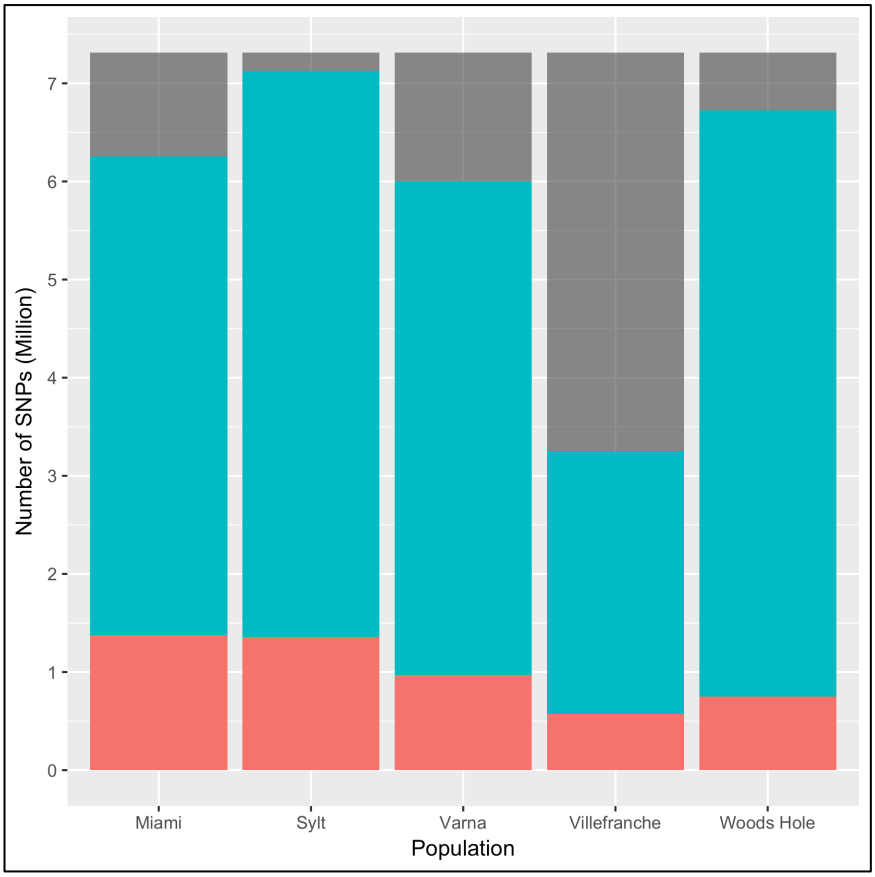

Figure 2: Population-specific SNP counts for the lenient (blue) and strict (red) sets. Grey bars show the total raw set on which filtering was performed.

3.2 Population Structure

3.2.1 Neighbour-joining Dendrogram

The neighbour-joining dendrogram (Figure 3) revealed primary structuring into two major haplogroups with Woods Hole and Sylt samples forming a "northern" group and Miami, Varna and Villefranche specimens comprising a "southern" group. Within the southern cluster, samples form well-supported monophyletic clades according to sampling location, *i.e.* every individual is more closely related to specimens from its own geographic location than to any other. Miami individuals are genetically more distant to both Varna and Villefranche samples which comprise a monophyly of their own. In contrast to the southern sites, northern samples display distinctly smaller branch lengths and weaker separation by geographic location. Whilst Woods Hole samples form a monophyly, Sylt specimens form a paraphyletic group indicating that some individuals are genetically closer to Woods Hole than to fellow Sylt specimens.

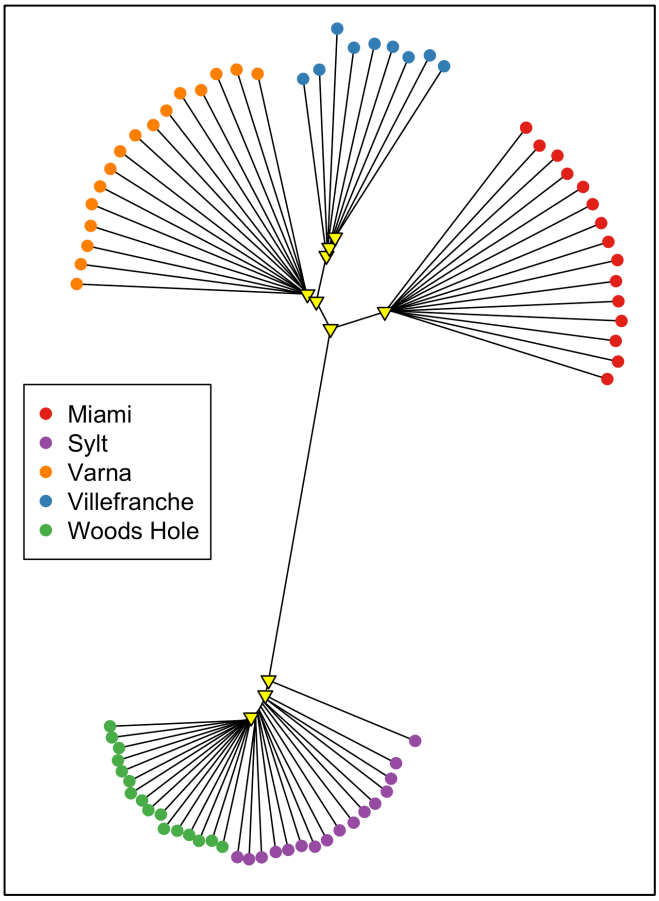

Figure 3: Neighbour-joining dendrogram of individual specimens color-coded by sampling location. Yellow triangles indicate nodes with >95% bootstrap support.

Finally, it should be noted that the relatively long terminal branches are most likely caused by the high number of markers used to generate the dendrogram. The inclusion of more markers inherently increases the number of sequence mismatches even between closely related individuals thus inflating genetic distance. Nonetheless, nodes generated from unstructured SNPs do not receive high bootstrap support and can be neglected, as is the case for most nodes at the base of the terminal branches.

3.2.2 Principal Component Analysis

The scree plot of explained variance indicates three principal components (PCs) to be responsible for 48.2% of the total genetic variation between all samples (Figure 4). Following Cattell's Rule, the inclusion of more PCs would not have increased information content significantly so lower order PCs were not considered further. Note that the first PC is responsible for over 35% of the explained variance alone suggesting strong primary structuring into two major haplogroups.

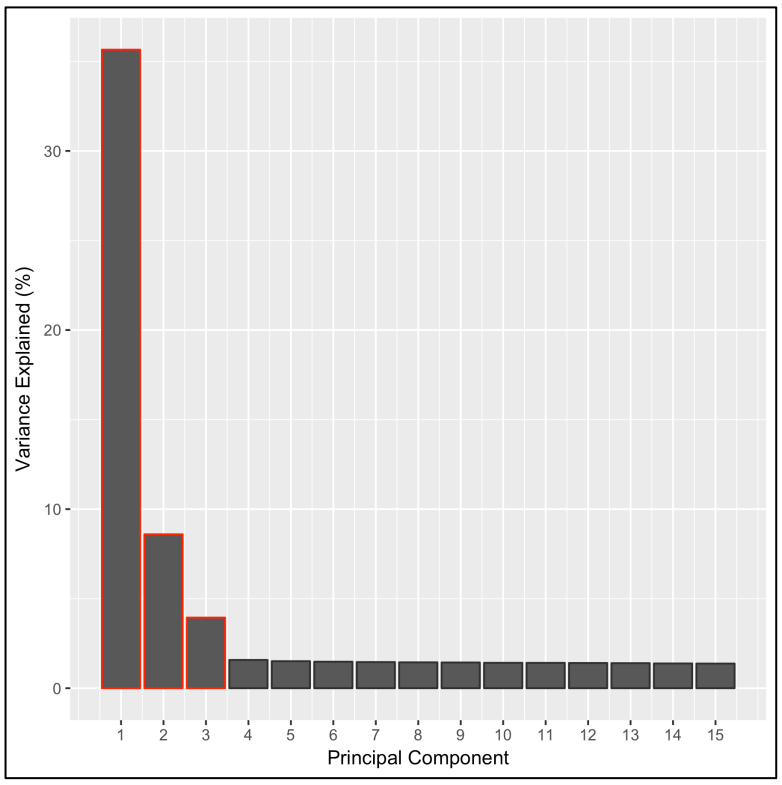

Figure 4: Scree plot of variance explained by the first 15 principal components in the global PCA. Red-highlighted PCs were retained as suggested by Cattell (1966).

The projection of samples onto the first three PCs (Figure 5) yields tight clusters corresponding to sampling locations implicating low intra- as compared to inter-population variance. Similarly to the neighbour-joining dendrogram, primary structuring occurs between the northern populations, Woods Hole and Sylt, and the southern populations, Miami, Varna and Villefranche along the first PC. The second PC further separates the southern group into the native Miami and the invasive Varna and Villefranche populations, with Varna specimens clustering slightly closer to Miami than Villefranche does. Finally, the third PC segregates Varna and Villefranche individuals with the degree of genomic differences decreasing with every order of principal component as illustrated by the scree plot.

Narrow peaks in the density distributions of Woods Hole along all PCs imply low within-population variance, possibly indicative of lower diversity. In contrast, more missing data in the under-sequenced Villefranche and, in part, Varna individuals may have caused clusters to elongate towards the origin leading to artificially broad clustering, most noticeable in the projection onto the first and second PC. Such trending towards the origin of the graph was mitigated through the use of the strict, highly complete, global SNP set albeit not completely eradicated. For the highly covered Sylt specimens, missing data does not explain the broader clustering. While Miami individuals clearly separate from the southern invasive populations along the second PC it should be noted that separation between Miami and Varna is greater than between Miami and Villefranche suggesting higher genomic similarity between the former two.

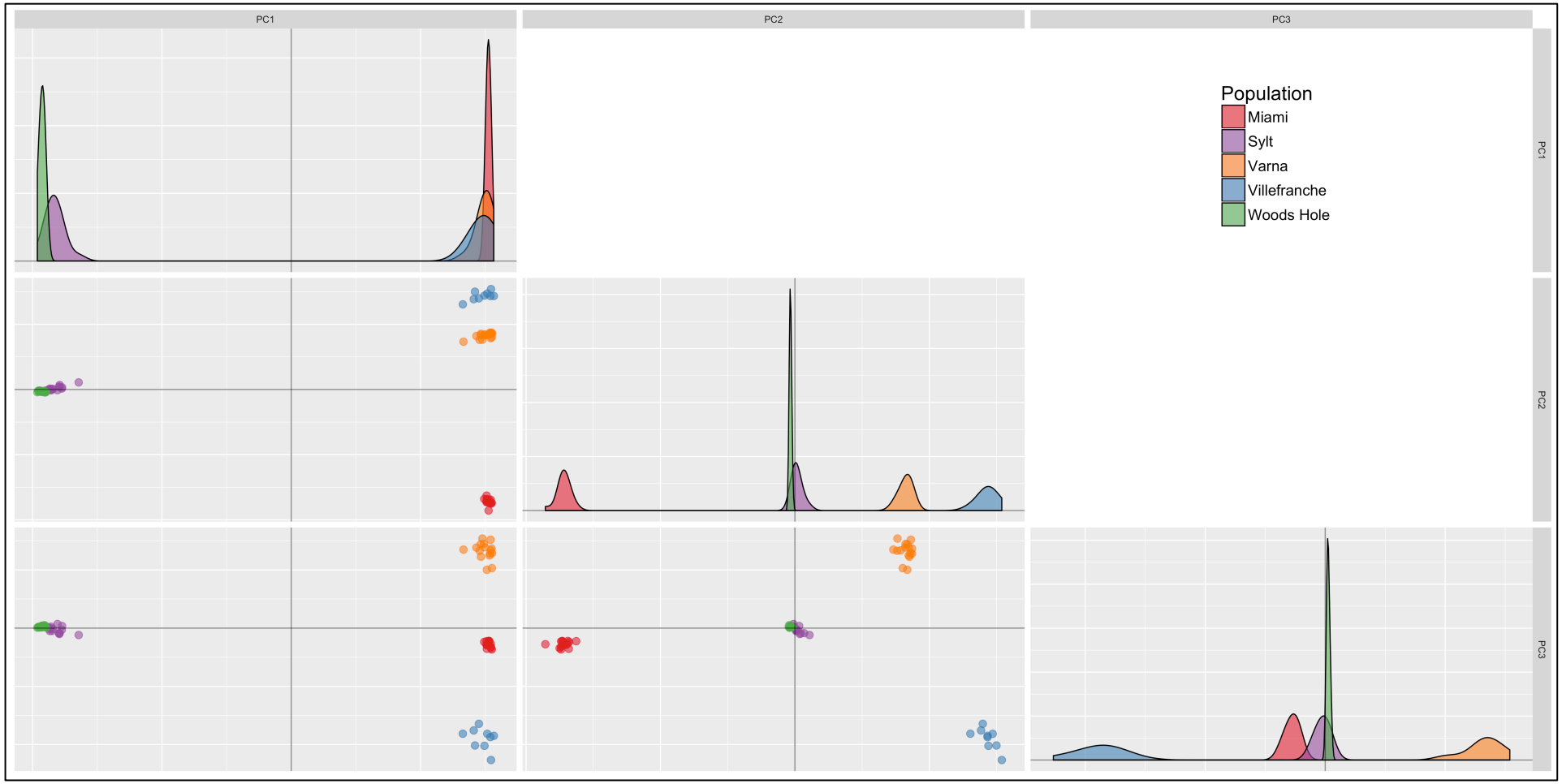

Figure 5: Projections onto the first three principal components of the global PCA. Graphs on the lower left display individual samples projected onto two PCs. Figures along the diagonal show density distributions along a single PC coloured by population.

The northern populations, Sylt and Woods Hole, did not resolve along any of the three PCs. While this overlap could be indicative of gene flow between the two populations, genomic divergence could still be present but at a much smaller scale than that observed at the global level. By conducting an independent PCA of the northern populations in isolation it was indeed possible to resolve Woods Hole and Sylt clusters though only to a minor degree. The scree plot of explained variances (Figure 6) reflects the lack of clear structure with the two major PCs only explaining 11.6% of the total variance. Lower-order PCs have eigenvalues of uniformly decreasing magnitude and their inclusion yielded no additional information. Some organisation into separate clusters is discernible in the projection plots (Figure 7), especially along the first PC, although this only accounts for 6.3% of the total variance. The second PC emphasises the variance within the Sylt population, which displays markedly loose clustering when compared to the tightly grouped Woods Hole individuals. As evidenced by the density distributions, the variance within the Sylt population is larger than the inter-population variance. In fact, some Sylt individuals project nearer to Woods Hole specimens than to fellow Sylt samples, in accordance with the paraphyletic grouping noted above.

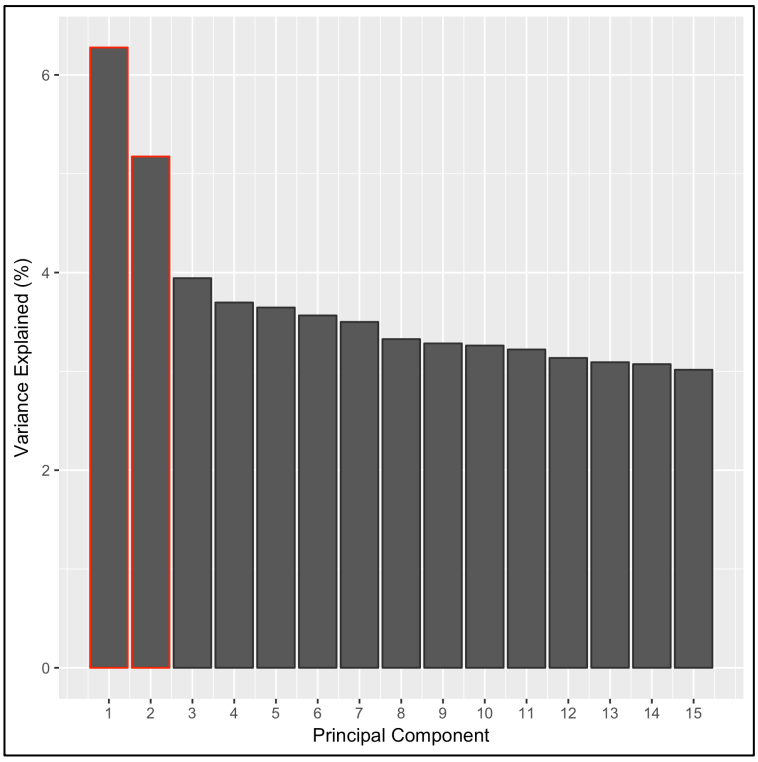

Figure 6: Scree plot of variance explained by the principal components in a PCA comprising only Woods Hole and Sylt samples. Projections were made onto the first two PCs (red).

Figure 7: PCA of the northern populations only. Lower left graph displays individual samples projected onto the first two PCs. Figures along the diagonal show density distributions along a single PC coloured by population.

To assess whether the principal components derived from the isolated, northern populations could also separate the southern populations, all Miami, Varna and Villefranche samples were projected onto the northern PCs after their construction (Figure 8). As expected, southern populations display compact clustering since PCs were not constructed to maximise their variance. Nonetheless, some separation along the first PC can be discerned though only to a minor degree (Supplement M). Differences in the projections of Sylt and Woods Hole samples between Figures 7 and 8 arise from the application of distinct SNP sets. The projection of southern samples required the usage of SNPs common to all populations whilst SNPs common to only Sylt and Woods Hole were sufficient for the independent, northern PCA shown in Figure 7. These differences in the input genotype matrix led to inconsistent projections although the qualitative pattern, as well as the loadings of the PCs, remained unchanged.

An independent PCA assessing the southern samples in isolation with subsequent projection of northern specimens was also performed (Supplement N). As expected, this analysis reproduced the second and third PC from the global PCA yielding no further information and confirming that, besides the split into northern and southern haplogroups, genomic divergence between the southern populations dominates global population structure.

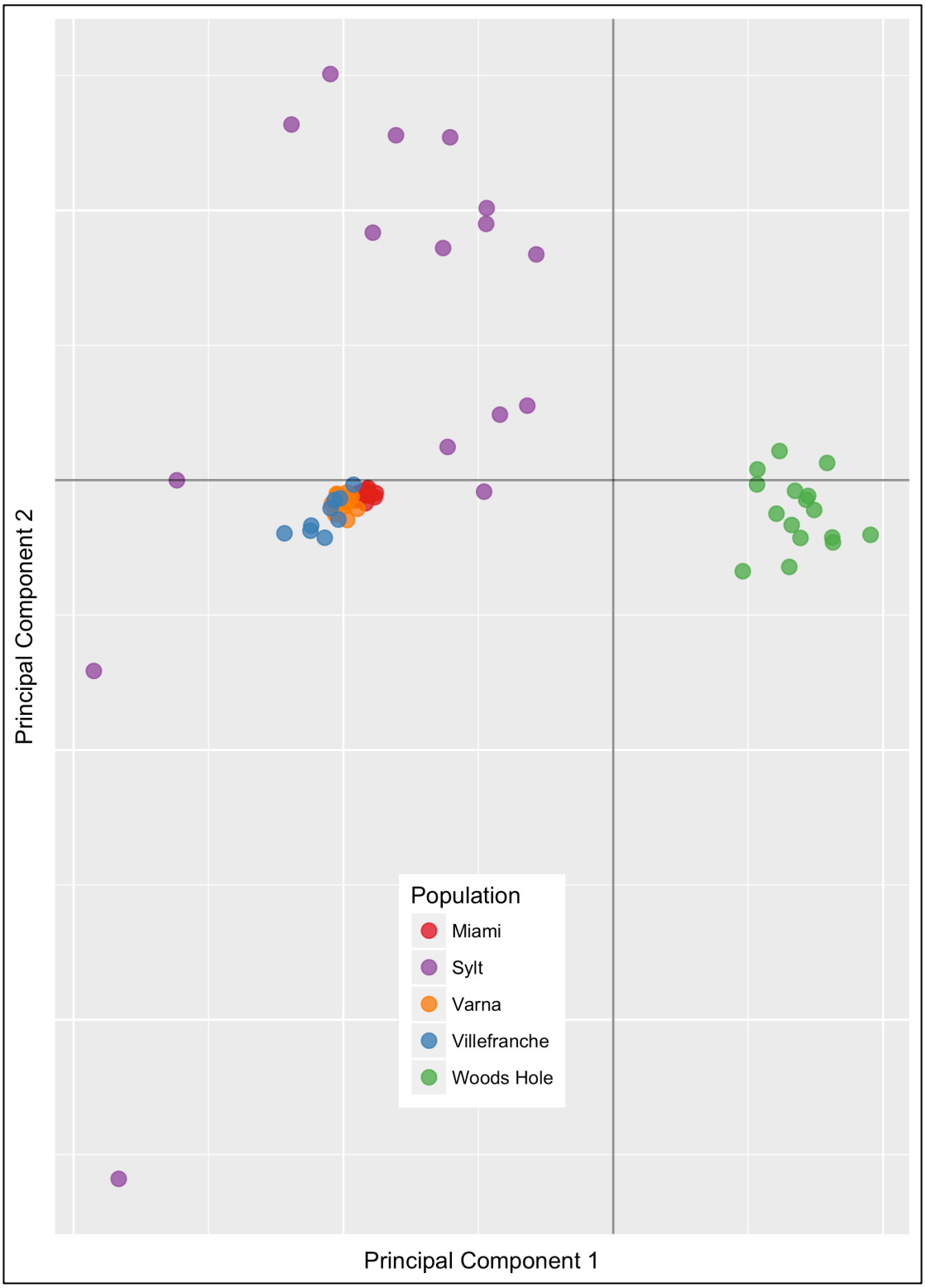

Figure 8: Southern populations projected onto the principal components independently derived from a PCA of the northern populations.

3.2.3 Hardy-Weinberg Equilibrium

The distribution of q-values resulting from the HWE exact test on all loci in the global, lenient SNP set is presented in Figure 9. 13.8% of all loci tested significantly at a false discovery rate (FDR) of α = 0.05. As shown in Figure 10, the vast majority (98.2%) of these loci fail due to a heterozygote deficit with only a minor amount in disequilibrium due to heterozygote excess. Reproduction through self-fertilisation might lead to some loss of heterozygosity but the strong population structure evidenced above is most likely to be the cause of this result.

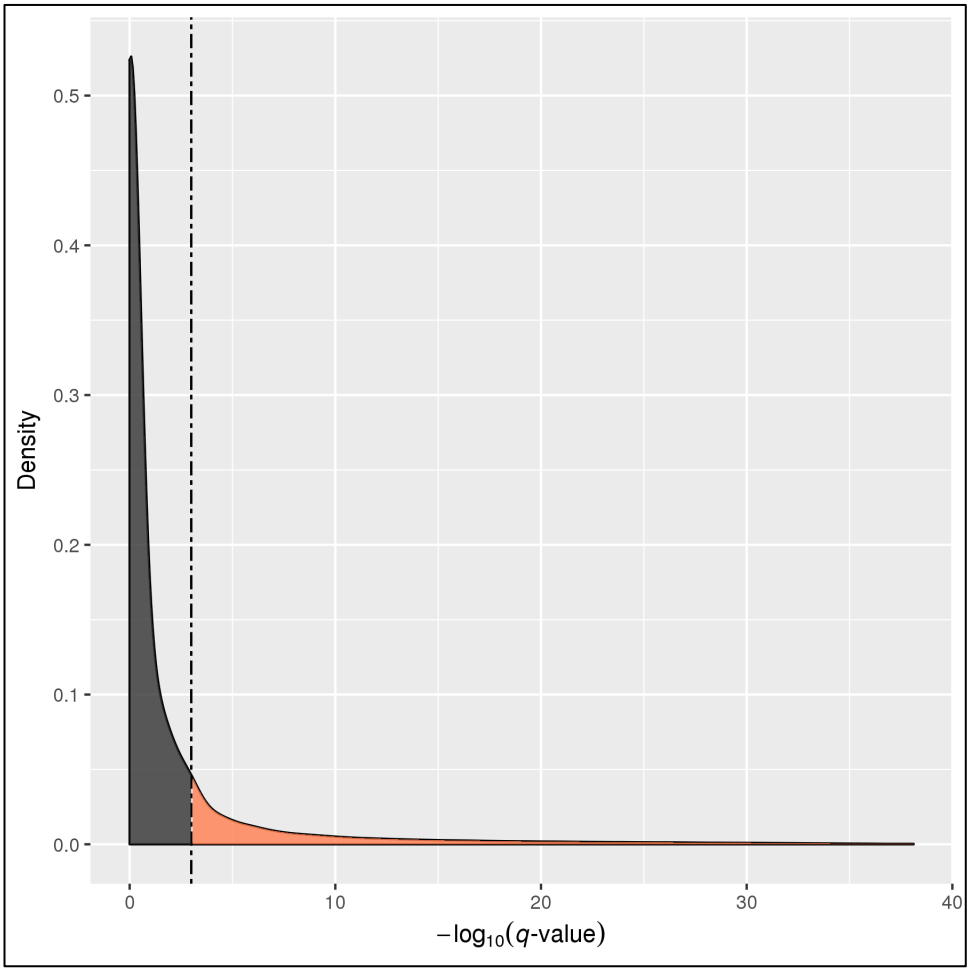

Figure 9: Distribution of q-values resulting from HWE exact tests on all loci of the global, lenient SNP set. The dashed line indicates the threshold of α = 0.05. The red-shaded area shows loci testing significantly for disequilibrium.

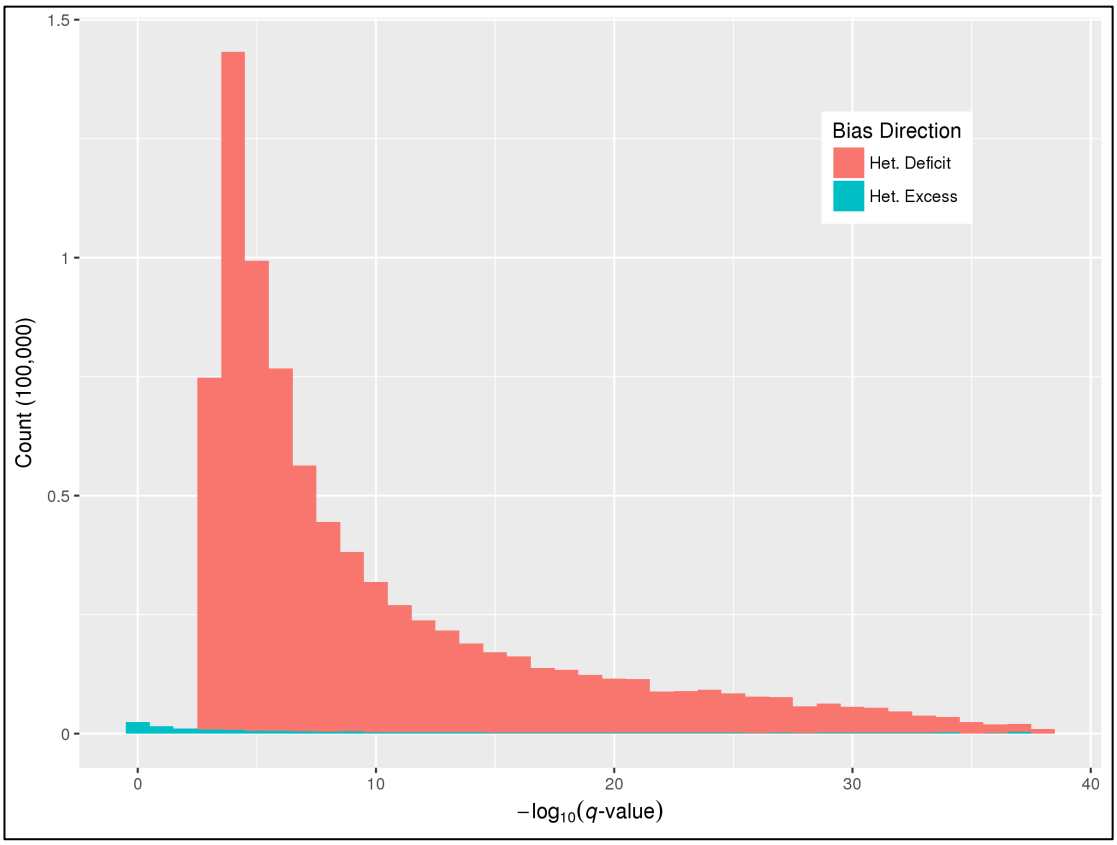

Figure 10: Histogram of *q*-values and directionality of loci failing the HWE exact test. Red bars indicate loci failing due to heterozygote deficit, blue bars due to heterozygote excess.

Conducting HWE exact tests with $\alpha = 0.05$ at the population level resulted in only 0.3% of loci displaying disequilibrium for the Sylt and Woods Hole populations and no loci testing significantly for Miami, Varna and Villefranche. All five populations were therefore deemed to be in HWE, justifying subsequent population-specific analyses requiring the assumption of HWE.

3.2.4 Global and Pairwise F_{ST}

Figure 11 shows the distributions of weighted and unweighted, global F_{ST} estimates calculated in overlapping windows of 10 Kb. The weighted F_{ST} estimate proved to be significantly higher than the unweighted estimate (paired *t*-test, $p<<0.05$) with a genome-wide mean of 0.332 as compared to 0.166 respectively. Genome-wide variance for the weighted F_{ST} was also higher at 0.013 compared to 0.002 for the unweighted estimate. Both estimates show remarkable intra-genomic variation with F_{ST} values of up to 0.898 in highly divergent regions and down to -0.020 in regions exhibiting heterozygote excess.

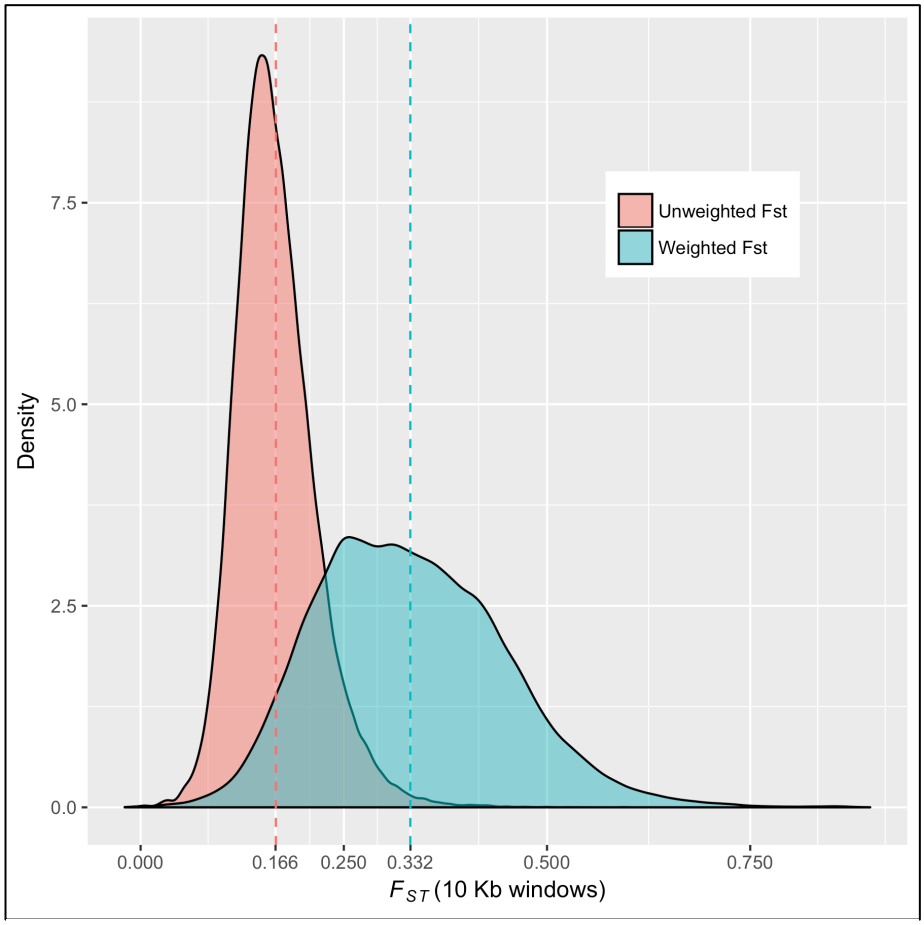

Figure 11: Distributions of weighted (blue) and unweighted (red) F_{ST} estimates calculated in overlapping windows of 10 Kb. Dashed lines indicate the respective mean values.

Pairwise, genome-wide F_{ST} estimates for all populations are presented in Table 2. Weighted means are consistently larger by a factor 2 yet both estimates qualitatively agree with the PCA as well as the NJ dendrogram. Sylt and Woods Hole populations show minimal genomic divergence with weighted (unweighted) fixation indices of 0.021 (0.016). Southern populations display higher F_{ST} values between themselves with Miami showing more similarity to Varna than to Villefranche specimens, exhibiting indices of 0.101 (0.058) as compared to 0.146 (0.084). Varna and Villefranche are the most related of the southern populations with F_{ST} values of 0.069 (0.038). Divergence between northern and southern populations is universally high ranging from 0.402 to 0.493 for the weighted and 0.203 and 0.317 for the unweighted estimates, with Woods Hole and Villefranche displaying the largest fixation indices. It should be noted that unweighted F_{ST} estimates are included for comparative purposes only. Weighted estimates are preferred, especially for large SNP sets with a high proportion of rare alleles where the unweighted method tends to underestimate the true F_{ST} (Bhatia *et al.*, 2013).

Table 2: Mean, pairwise F_{ST} estimates calculated in non-overlapping windows of 10 Kb. Upper diagonal contains weighted, the lower diagonal unweighted means. All values tested significantly different from zero (Supplement O).

	Woods Hole	***Sylt***	***Miami***	***Varna***	***Villefranche***
Woods Hole	-	*0.021*	*0.434*	*0.427*	*0.493*
Sylt	*0.016*	-	*0.410*	*0.402*	*0.459*
Miami	*0.220*	*0.210*	-	*0.101*	*0.146*
Varna	*0.212*	*0.203*	*0.058*	-	*0.069*
Villefranche	*0.317*	*0.299*	*0.084*	*0.038*	-

3.3 Genomic Diversity

3.3.1 Multilocus Heterozygosity

To minimise ascertainment bias in the calculation of individual multilocus heterozygosity (IMLH) as well as population-wide heterozygosity (H_S), groups were randomly subsampled to the largest common sample size of $n = 9$. Since Villefranche was the limiting population it was not subsampled and all specimens were retained in the analysis. A boxplot of IMLH and H_S is shown in Figure 12. Naturally, the two indicators are closely linked with H_S falling close the median IMLH for all populations. Miami and Varna specimens exhibited the lowest H_S of 0.263 and 0.260 respectively. Varna specimens showed higher variance in IMLH, though means are not significantly different (Supplement P). The remaining populations display significantly higher values both for IMLH and H_S with Woods Hole exhibiting the highest H_S overall of 0.306. Sylt and Villefranche present similar H_S of 0.294 and 0.293 respectively. Whilst dissimilarities are apparent, absolute differences are small with all estimates of population-wide heterozygosity falling into a range of less than 0.05.

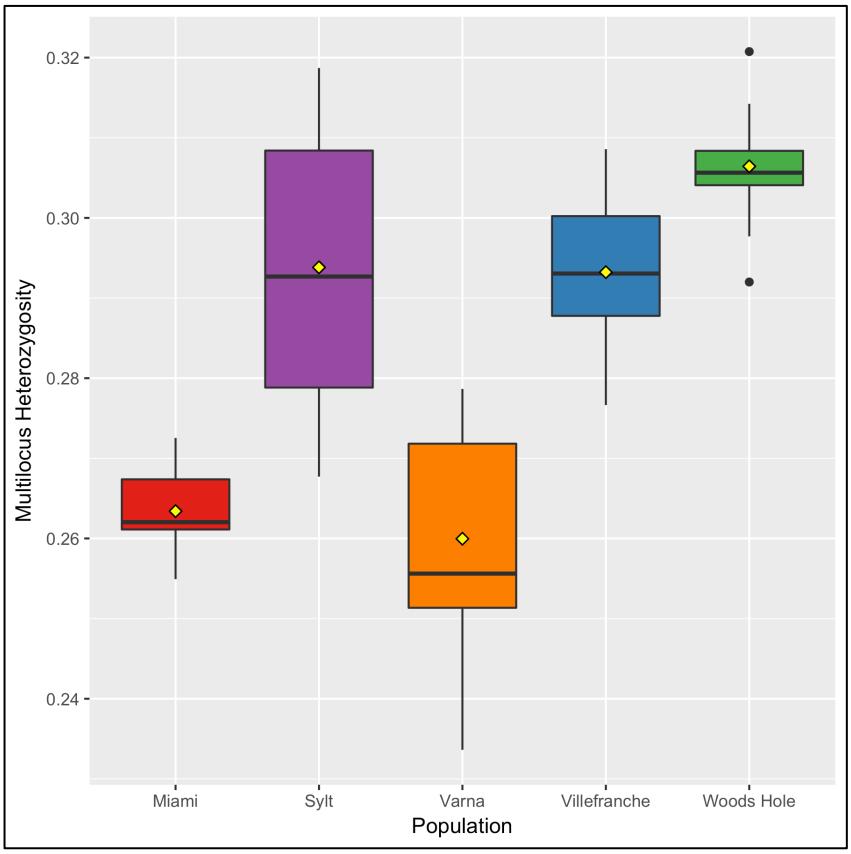

Figure 12: Boxplot of IMLH grouped by population. Yellow squares denote the population-level, mean heterozygosity, H_S.

3.3.2 Watterson's Theta

The genome-wide distributions of Watterson's theta, θ_W, in windows of 10 Kb are shown in Figure 13. Miami and Varna show highly similar distributions with insignificantly different means of 6.48 $x10^{-3}$ and 6.46 $x10^{-3}$ respectively (Supplement Q). Woods Hole and Sylt also display high similarity in the shape of their distributions though Woods Hole has a significantly lower mean of 2.77 $x10^{-3}$ compared to the Sylt's 3.70 $x10^{-3}$. Villefranche specimens exhibit the lowest variance in θ_W as well as the lowest mean of 2.35 $x10^{-3}$. Miami and Varna populations show a distinctly more heterogeneous distribution of θ_W than Woods Hole and Sylt, with a higher proportion of divergent regions. In contrast, peaks at the lower end in the distributions of the northern samples indicate multiple conserved regions with a low number of segregating sites. The proportion of these low-diversity regions is higher than expected and suggestive of stronger haplotype structure in the northern populations.

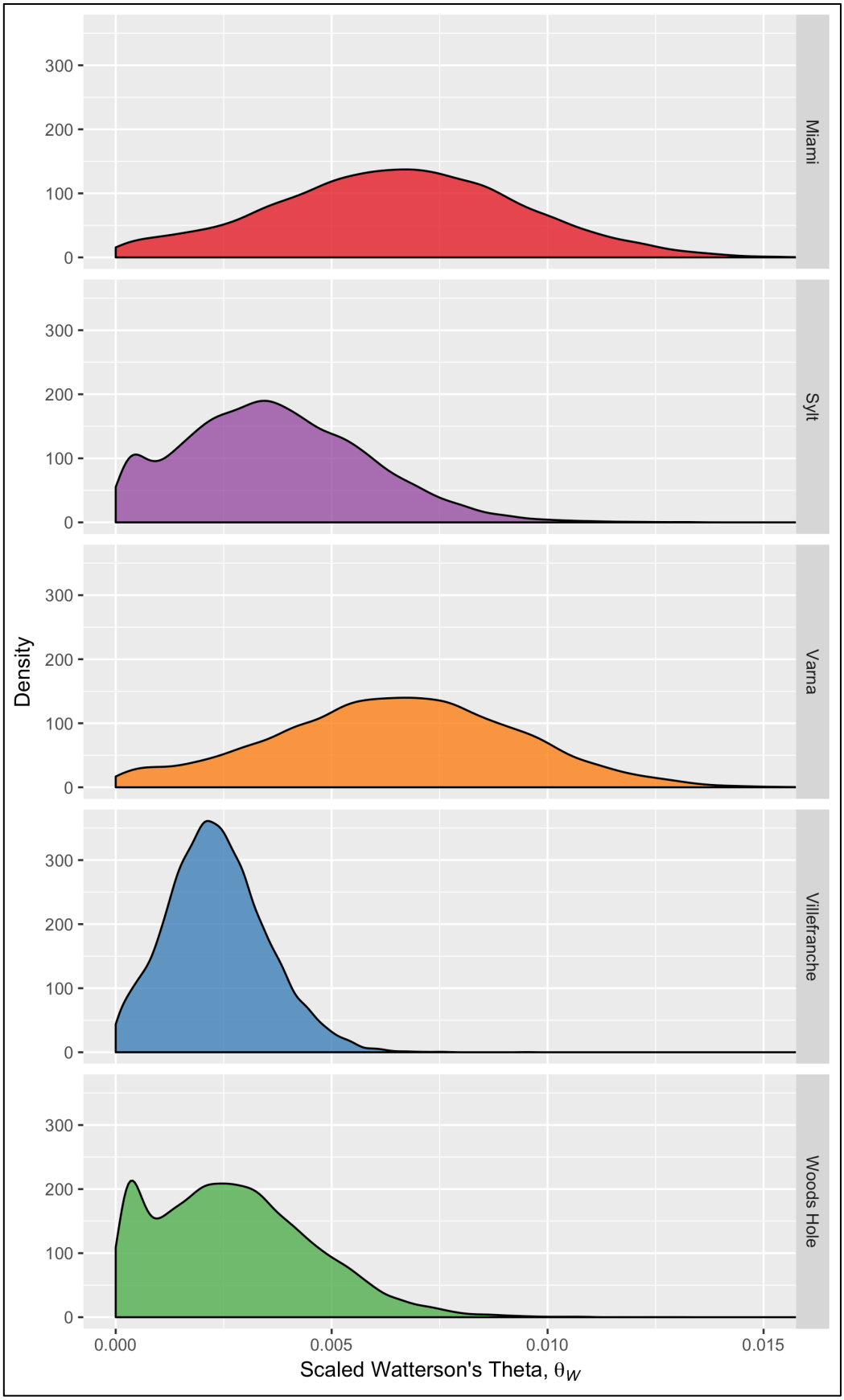

Figure 13: Genome-wide distribution of Watterson's theta, θ_W, in non-overlapping windows of 10 Kb. θ_W is scaled to the window size and hence a per-base statistic.

3.3.3 Nucleotide Diversity

Since nucleotide diversity, π, and θ_W are two alternative estimates of polymorphism in a population, it is not surprising that the genome-wide distributions are highly comparable (Figure 14). Again, Miami and Varna display similarly broad distributions with the highest mean nucleotide diversities of 5.79 $x10^{-3}$ and 5.39 $x10^{-3}$ respectively (Supplement R). Villefranche exhibits the most homogeneous nucleotide diversity across the genome as well as the lowest mean of 2.13 $x10^{-3}$. Woods Hole and Sylt show intermediate diversities of 3.13 $x10^{-3}$ and 4.26 $x10^{-3}$. Once again, their distributions depart from normality due to an excess of low-diversity genomic regions leading to a peak at lower values of π. A noteworthy difference to θ_W is that Woods Hole and Sylt exhibit remarkably higher variance in π, especially Sylt displays nucleotide diversity on par with Miami and Varna at the higher end of its distribution.

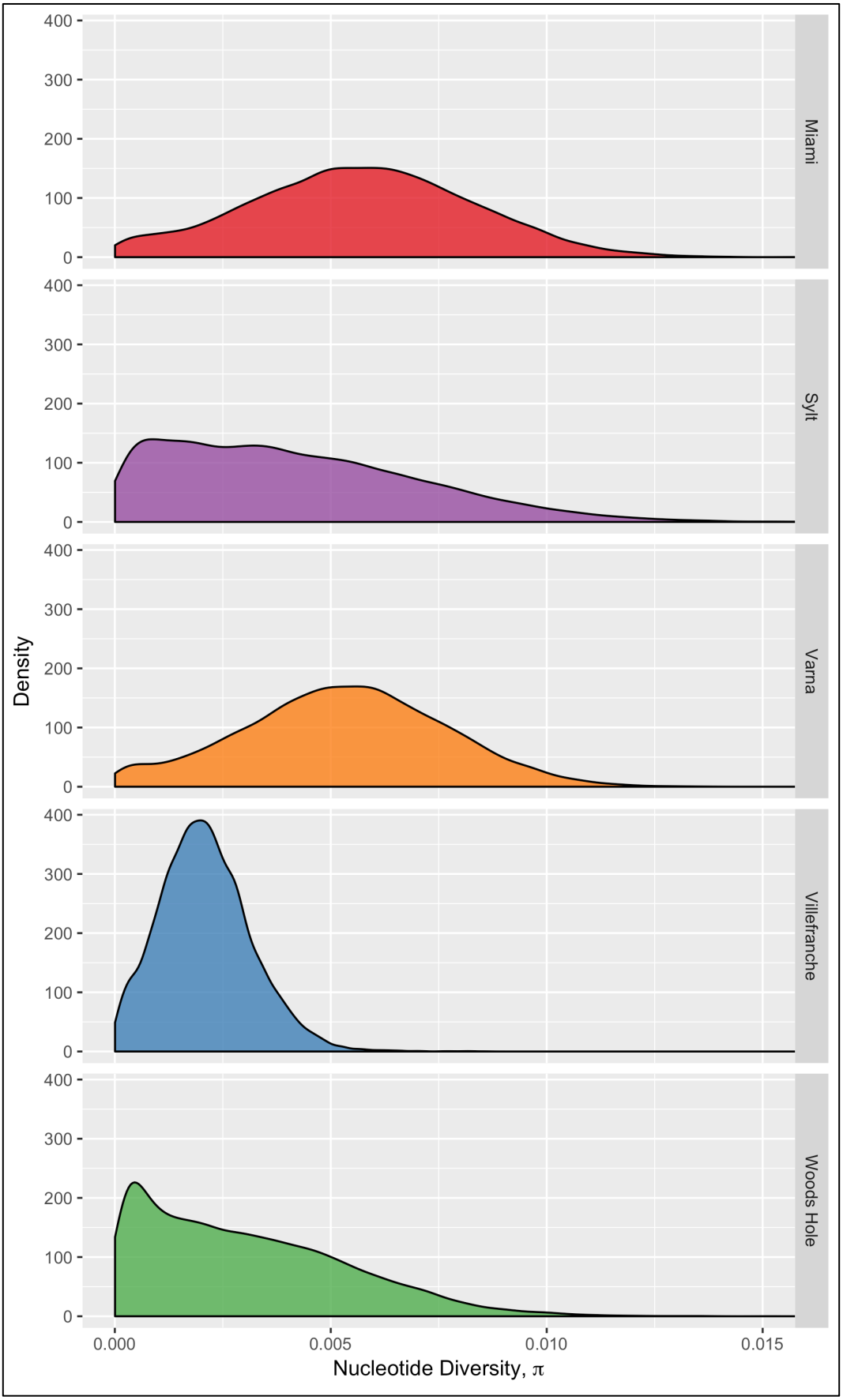

Figure 14: Population-specific distributions of nucleotide diversity, π, calculated in non-overlapping windows of 10 Kb. Nucleotide diversity is scaled to the window since and is thus a per-base statistic.

3.3.4 Coefficients of Inbreeding

Estimates of individual inbreeding coefficients, *F*, as well as the population-level statistic, F_{IS}, are plotted for each population (Figure 15). As expected, the estimates of F_{IS} fall close to the median *F* for all populations. Both northern locations exhibit negative inbreeding coefficients with no significant difference between them, though only Woods Hole shows significant difference from zero (Supplement S). Southern populations also show no disparity between themselves with only Miami being significantly different from zero. Miami displays the highest F_{IS} of 0.013 whilst Woods Hole takes the lowest value at -0.033. Although individual outliers with higher levels of inbreeding (*F* = 0.099) or outcrossing (*F* = -0.128) exist, mean F_{IS} for each populations is low enough to be deemed negligible.

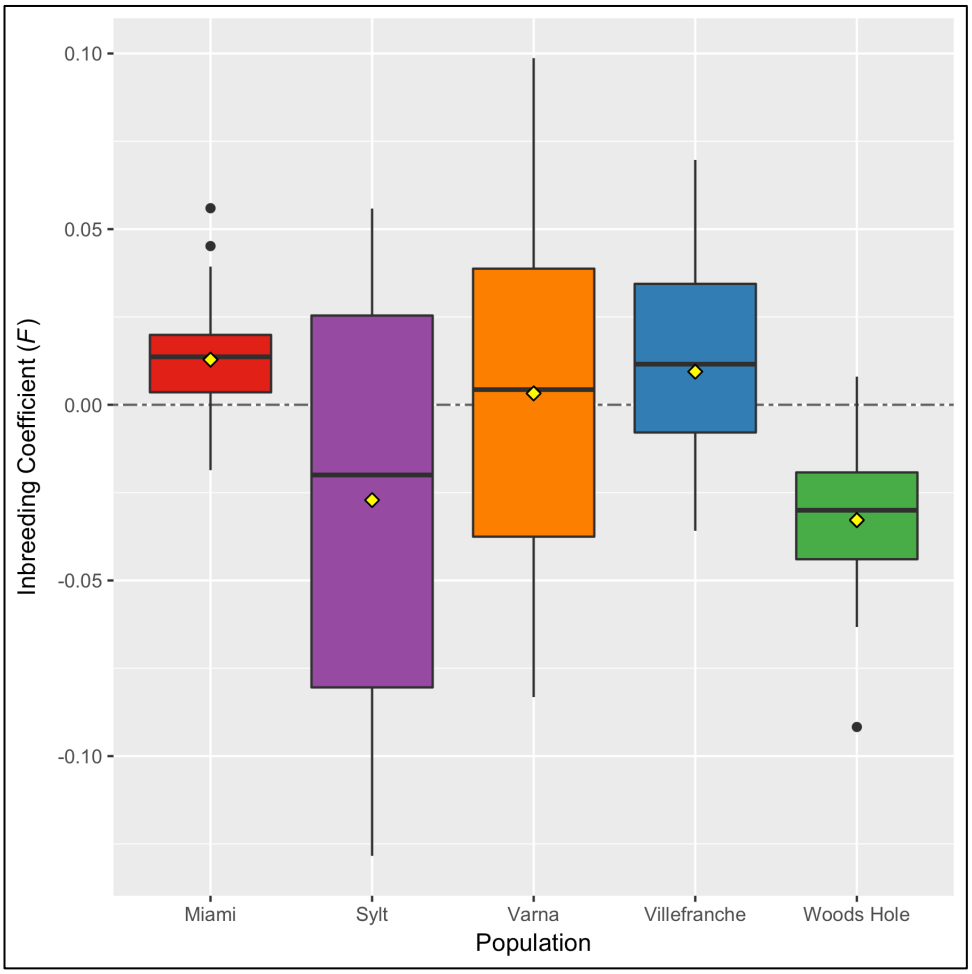

Figure 15: Boxplot of individual inbreeding coefficients, *F*, grouped by population. Yellow squares denote the population-level estimate, F_{IS}, after Weir and Cockerham (1984).

3.4 Demographic Indicators

3.4.1 Allele Frequency Spectrum

Prior to the construction of the folded allele frequency spectrum (AFS), populations were randomly subsampled to the largest common sample size of $n = 9$. Population-specific minor allele counts summed and normalised across all loci are given in Figure 16. In a sample of 9 diploid individuals, the minor allele can occur between 1 and 8 times at each locus, elsewhere it is undefined.

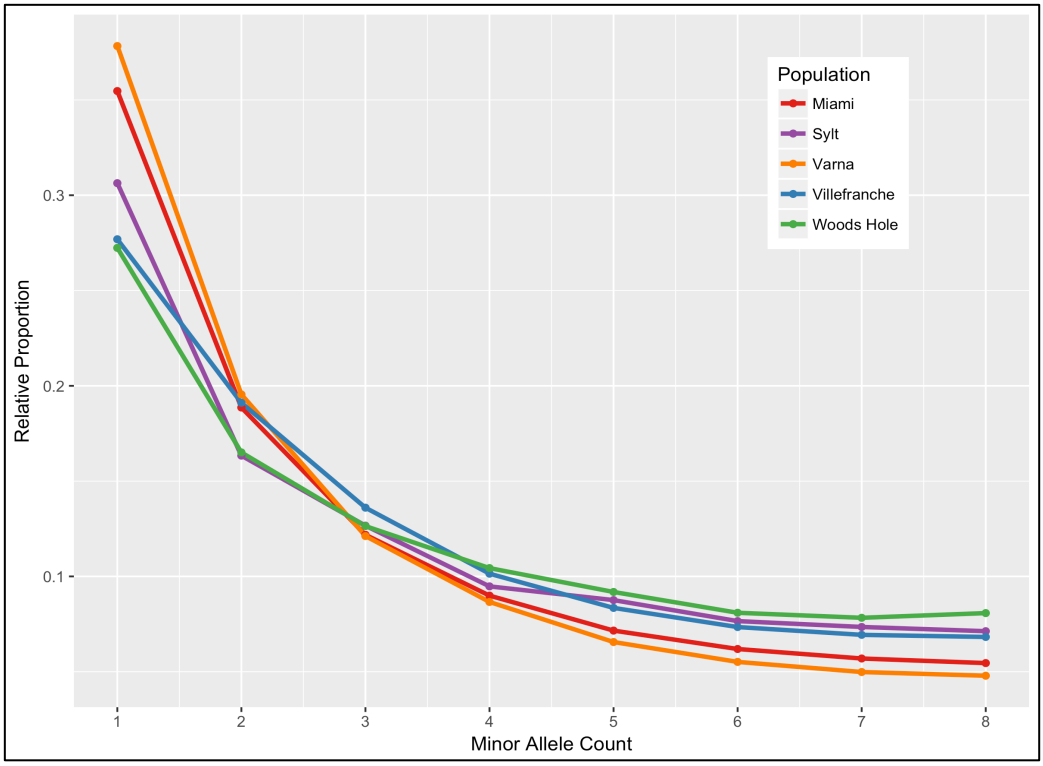

Figure 16: Folded allele frequency spectra coloured by population. The horizontal axis contains the number of minor alleles counted at a locus. The vertical axis describes the relative proportion of loci with a given minor allele count.

All populations display the typical shape of a folded AFS; high proportions of rare SNPs with low minor allele counts and fewer loci at intermediate allele frequencies, nonetheless inter-population differences are apparent (Supplement U). Miami and Varna exhibit the steepest gradient due to a higher proportion of rare SNPs and fewer loci with balanced allele frequencies. The northern populations, Woods Hole and Sylt, show the opposite pattern, with a slightly more uniform distribution of minor allele frequencies. Villefranche's AFS resembles that of the northern populations, displaying relatively few rare variants and a higher proportion of high-frequency SNPs.

3.4.2 Tajima's D

The genome-wide distributions of Tajima's *D*, calculated for each population in non-overlapping windows of 10 Kb are given in Figure 17. As Tajima's *D* is essentially a measure of the discrepancy between the two polymorphism estimators, θ_W and π, one can expect *D* to fall close to zero for genomic regions with similar θ_W and π. Again, Miami and Varna show almost identical distributions, approximately centred on but still significantly different from zero, with means of -0.088 and -0.142 respectively (Supplement T). Variance in *D* is similar and comparatively low for both populations. In contrast, the distributions of Sylt and Woods Hole are remarkably broader indicating more heterogeneity across the genome and a large proportion of windows exhibiting extreme values of *D*. Mean values of 0.145 and 0.389 are also significantly different from and close to zero, but positive as compared to Miami and Varna. Villefranche takes a mean of positive 0.357 with a relatively low variance, similar to those of Miami and Varna.

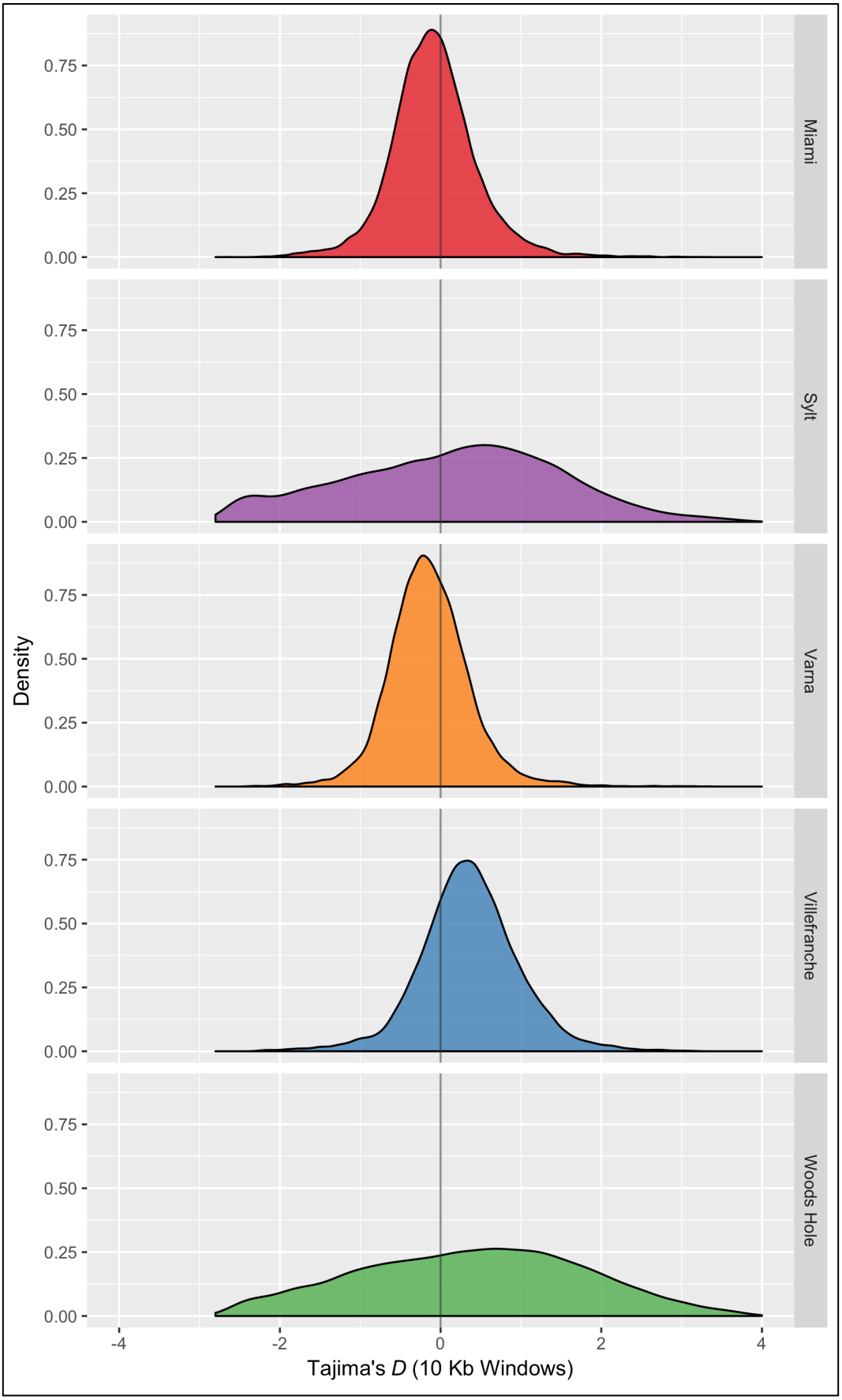

Figure 17: Distributions of Tajima's *D* in the genomes of five populations, calculated in non-overlapping windows of 10 Kb. The grey line indicates the neutral expectation of zero.

3.4.3 Linkage Disequilibrium and Effective Population Size

The graph of linkage disequilibrium (LD) decay presented distinct signatures for all five populations (Figure 18). Woods Hole and Sylt displayed the highest short-range LD with the correlation coefficient, r^2, returning values above 0.4 for loci less than 1Kb apart. LD decayed more rapidly with distance in the Woods Hole population with a half-decay distance of 9 Kb compared to Sylt's 13 Kb. Woods Hole also reached lower long-range r^2 with a mean of 0.010 above 300 Kb, in contrast with Sylt's 0.025. Southern populations exhibited a dissimilar pattern with considerably lower short-range LD. Miami and Varna present maximum r^2-values around 0.1 for loci less than 1 Kb apart. LD decay is more pronounced in Miami specimens, with a half-decay distance of 8 Kb, than in Varna individuals, 12 Kb. However, both Miami and Varna reach similar mean r^2-values at distances above 300 Kb (0.008 and 0.010). The Villefranche LD decay signature is akin to those of the other southern populations, yet upshifted both at short and long-range by approximately 0.05. This gives Villefranche the highest long-range LD of 0.033 and the highest half-decay distance of 55 Kb.

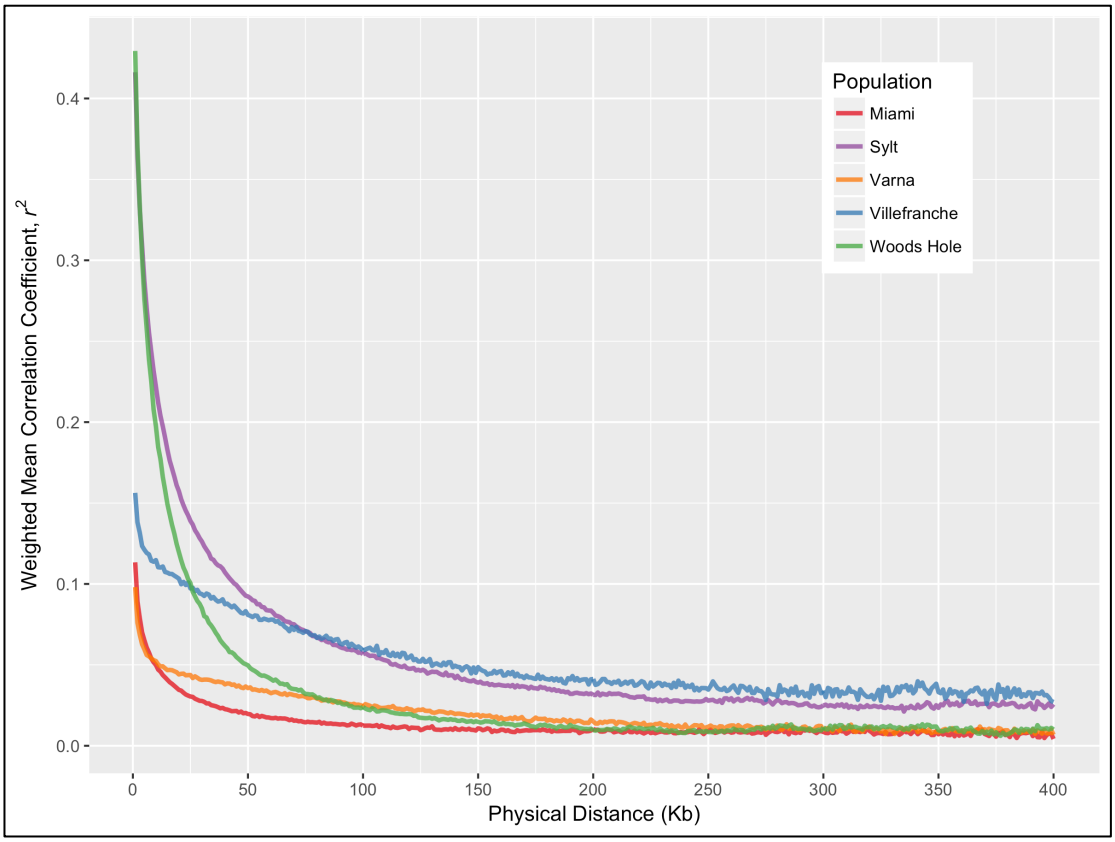

Figure 18: Decay of linkage disequilibrium as a function of physical distance in Kb. The vertical axis represents the weighted, mean correlation coefficient between genotypes, r^2.

Dynamic estimates of effective population size, N_e, computed from LD data shown above, are presented in Figure 19. Estimates only extend 150 generations into the past due to the scarcity of point estimates in deeper time and the focus of this study on the recent demographic history. Still, a plot of estimates extending further into the past can be found in Supplement V.

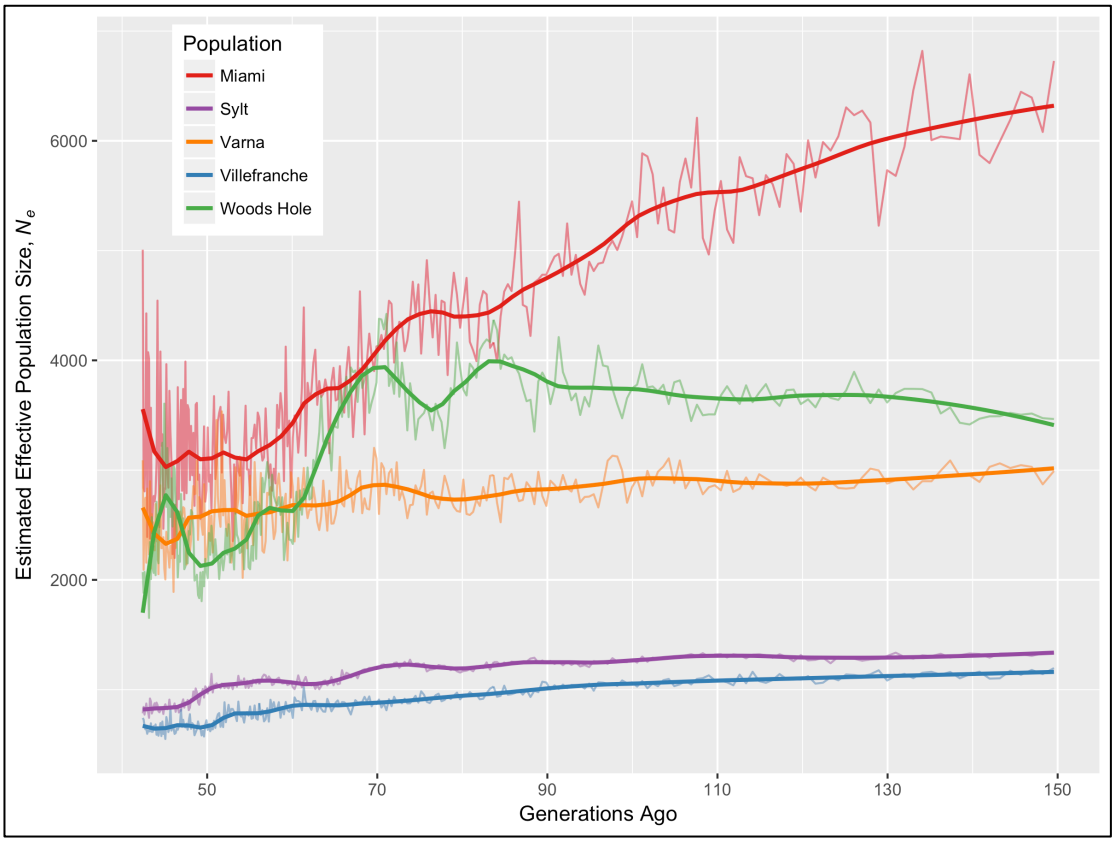

Figure 19: Estimated effective population size, N_e, as a function of generations in the past. Estimates were generated from unphased LD data using the method of Corbin *et al.* (2012). Lightly shaded lines link the point estimates of N_e, darker lines represent the running mean of the point estimates.

As a result of the distinctive LD decay patterns, the estimates of effective population size also display diverse signatures. The Miami population shows a clear decline in size during the recent past, dropping from over 6000 effective breeders to approximately 3000. Nonetheless, Miami shows signs of rapid expansion during the past 25 generations recovering to over 3500 effective breeders. Varna exhibits a mild decline in N_e from 3000 to 2500 but, like Miami, shows a brisk increase in effective population size in the very recent past. Woods Hole displays stable N_e of just under 4000 up until 70 generations in the past from when it seems to have undergone a steady decline in size down to almost 2000. After a brief period of rapid recovery the Woods Hole population seems to

be declining drastically throughout the most recent 25 generations. The invasive populations Sylt and Villefranche present the most stable as well as the lowest estimates of effective population size with minimal fluctuations in point estimates. Both show a gradual decline in size with the most recent estimates of N_e coming to 900 for Sylt and 800 for Villefranche. However, for these young invasive populations, estimation of N_e for time points prior to their establishment is meaningless and extended projections into the past should be ignored. In fact, the model also assumes constant linear growth (Hayes *et al.*, 2003), which is certainly not the case during *M. leidyi* blooms. Estimation of effective population size based on unphased LD data therefore requires caution and results should be interpreted qualitatively rather than quantitatively.

4. Discussion

4.1 Organisation of the *M. leidyi* invasion system

By employing a high-density SNP set sourced from whole-genome re-sequencing of five geographically separate populations, a robust depiction of the *Mnemiopsis leidyi* population structure in its native and invasive range was attained. The results of the neighbour-joining dendrogram, pairwise F_{ST} estimates and principal component analysis were all in agreement, mostly supporting previous findings. Whilst these earlier studies were able to make similar conclusions using comparatively limited marker sets (Reusch *et al.*, 2010; Ghabooli *et al.*, 2011, 2013; Bolte *et al.*, 2013; Bayha *et al.*, 2015), whole-genome sequencing has allowed for a more vigorous approach, elucidating certain unresolved aspects in the process. This is especially true of a marine species for which linkage disequilibrium is quickly reduced through panmixia in large populations, and the inference of genome-wide attributes using only a small set of markers is problematic (De Wit and Palumbi, 2013; Benestan *et al.*, 2015; Tepolt, 2015). Here, the application of a high-density marker set led to tight clustering of samples by geographic location with no apparent outliers indicative of low intra- compared to inter-population variance. In addition, all sampling locations were shown to be in Hardy-Weinberg equilibrium, suggestive of local panmixia and no further sub-structure. This allowed each location to be treated as a single, geographically isolated population with regard to all others and simplified comparisons among them.

4.1.1 The North-South Divide

Population structure proved to be hierarchical with the global PCA being most representative of this. The first principal component demonstrated the most evident genomic divergence to exist between northern and southern specimens, supported by universally high pairwise F_{ST} values approaching 0.5. Joint clustering of Woods Hole with Sylt as well as Miami with Varna and Villefranche clearly confirmed the source regions of the three invasive

populations. A strong north-south divide, and consequential assignment of source regions, has previously been shown (Reusch *et al.*, 2010; Ghabooli *et al.*, 2011, 2013; Bolte *et al.*, 2013; Bayha *et al.*, 2015) but never at the whole-genome level where it accounts for over a third of the total genomic variance. This divergence into two haplogroups seems to be independent of any recent invasion dynamics given that it is present in the native populations. Whilst Bayha *et al.* (2015) have identified the likely reason for the persistence of this divide, the origins remain unknown. For the purpose of this study, the north-south divide provides a useful backdrop for the comparison of two spatially and temporally independent invasion events by the same organism.

4.1.2 Black Sea Primary Invasion

The next structural level is given by the second principal component of the global PCA, which resolved the native Miami population from its invasive counterparts, Varna and Villefranche. Highly supported branching in the neighbour-joining dendrogram as well as significant pairwise F_{ST} values between Miami and Varna and Miami and Villefranche support these results. *M. leidyi* was first documented in Sudak Bay in 1982 (Pereladov, 1988), thought to have invaded in the ballast water of shipping vessels frequenting between major ports in the Gulf of Mexico and the Black Sea (Bayha *et al.*, 2015). Whilst large-scale gene flow might have occurred during these initial stages of invasion, non-zero fixation indices and highly resolved clustering suggest this is currently not the case. Given the time period of over 35 years since first establishment, substantial genetic distance between Miami and Varna appears to be the result of allopatric divergence via genetic drift. In addition, extremely disparate environmental conditions between the source and invasive region might have accelerated this divergence. For instance, native populations of *M. leidyi* around the Floridian Peninsula are subject to salinities of over 35 PSU (Kelble *et al.*, 2007) whereas the invasive Varna population regularly experiences salinities of 18 PSU in the western Black Sea (Stanev, 1990). Whilst the adult life-stage of the osmoconforming ctenophore is extremely tolerant to a wide range of salinities,

egg-production rates are severely affected (Jaspers, Møller and Kiørboe, 2011). It is therefore plausible that such extreme environmental differences may have led to strong selection on reproductive output under stress, which in turn would have accelerated divergence.

4.1.3 Mediterranean Range Expansion

Genomic distance between Miami and Villefranche proved to be higher than between Miami and Varna, evidenced by larger separation along the second PC, higher mean F_{ST} values and longer branch lengths in the dendrogram. This stands as strong evidence in favour of the stepping-stone hypothesis suggested by Bolte *et al.* (2013) who postulated the Western Mediterranean *M. leidyi* population to be the result of a range expansion out of the Black Sea. Elevated drift due to smaller effective population sizes during a range expansion may have caused Villefranche to diverge at a faster rate with respect to Miami than Black Sea populations. The hypothesis proposed by Ghabooli *et al.* (2013), who suggested the Western Mediterranean population to be the result of admixture between Black Sea and Gulf of Mexico populations, seems improbable in the light of the results presented here. Direct gene flow from Miami into the Western Mediterranean would require Villefranche to cluster between Varna and Miami and exhibit lower fixation indices. This is not the case and a pure range expansion from the Black Sea seems to be the most probable scenario. It should be noted, however, that sample sizes for Western Mediterranean populations were higher in previous investigations and incomplete sampling may have led to the omission of admixed individuals in this study. Nonetheless, low variance and HWE between specimens suggest that the existence of another haplogroup in the vicinity of Villefranche is unlikely.

The third level in structural hierarchy separated Varna from Villefranche with low, yet significant pairwise F_{ST} and bootstrap supported branching into two monophylies. Whilst both populations were already resolved along the second PC the third axis demonstrates genomic divergence orthogonal to Miami specimens. In other words, Varna and Villefranche are divergent for a set of SNPs that shows no pattern in Miami samples. This provides further evidence in favour of the range expansion from the Black Sea into the Mediterranean, as Miami specimens should display structure in this dimension if admixture was indeed occurring. *M. leidyi* was first recorded in the North-western Mediterranean in 2009 (Boero *et al.*, 2009; Fuentes *et al.*, 2009), over 20 years after its establishment in the Black Sea (Vinogradov *et al.*, 1989). Analogous to the primary invasion wave, the secondary step into the Mediterranean likely required adjustment to vastly dissimilar conditions. Divergence between Mediterranean and Black Sea populations due to geographic isolation and selection on *e.g.* salinities upwards of 39 PSU in the Eastern Mediterranean (Poulos, Drakopoulos and Collins, 1997) seems likely. In fact, the comparable oceanographic regimes found around the Floridian Peninsula and the Mediterranean beg the question whether the secondary invasion was made possible by the retention of pre-adapted alleles, possibly at low-frequencies, during establishment in the Black Sea. These pre-adapted alleles may then have been recycled during the expansion into the Mediterranean via a process similar to that found in threespine stickleback freshwater adaptation (Bell and Aguirre, 2013). However, the resurgence of alleles that are prevalent in Miami individuals would have resulted in higher genomic similarity between Mediterranean and Floridian samples, which is not the case.

A plausible scenario is that *M. leidyi* remained in the Black Sea for a substantial period of time prior to colonising the Mediterranean. Surface water outflow through the Sea of Marmara (Besiktepe *et al.*, 1994) may have seeded sink populations in the Aegean Sea where *M. leidyi* was first sighted in 1993 (Kideys and Niermann, 1994). However, these populations never fully established and significant *M. leidyi* blooms were not reported until 2009 (Fuentes *et al.*, 2009; Galil, Kress and Shiganova, 2009). Such a time lag of 20

years within the Black Sea would have eroded any pre-adaptations to Mediterranean conditions through drift and strong selection for the Black Sea environment. This would then require an original evolutionary path in order to invade the Mediterranean. The fact that Varna and Villefranche diverge orthogonally to Miami along polymorphisms that display no pattern in the native range suggests that Mediterranean invaders have found a novel path indeed.

4.1.4 The Young North Sea Invasion

The final, most vague, structural level occurred between the northern native and invasive population. While it was possible to resolve Woods Hole and Sylt into separate clusters via a dedicated PCA, genomic divergence was minimal relative to the other populations, leading to joint clustering in the global analysis. High genomic similarity was confirmed by an extremely low pairwise F_{ST} and paraphyletic branching in the NJ dendrogram. In fact, genomic distance between the two populations is such that certain Sylt specimens display higher similarity with Woods Hole individuals than with samples taken from the same location *i.e.* intra-population exceeds inter-populations variance. High relatedness between individuals from the North-western Atlantic and the North Sea had been demonstrated before, yet observing such similarity at the whole-genome scale for planktonic organisms either side of the Atlantic Ocean is remarkable. Compared to the southern system the northern invasion wave is relatively young with *M. leidyi* first reported in the North Sea in 2006 (Faasse and Bayha, 2006). Given *M. leidyi*'s short generation time of two weeks (Baker and Reeve, 1974), a small founding population in the North Sea should have experienced some drift during the past 10 years. The fact that Sylt shows little genomic divergence to Woods Hole after a decade of geographic isolation is suggestive of two scenarios. Either, i) primary inoculates were large, *M. leidyi* quickly established a sizeable population and the effect of genetic drift was weak from the onset, or ii) there continues to be sustained gene flow between the North-western Atlantic and the North Sea. These hypotheses are not mutually exclusive and equally troubling with regard to Northern Europe's susceptibility towards planktonic marine

invaders. Europe's most frequented ports: Rotterdam, NL, Antwerp, BE and Hamburg, DE (Eurostat, 2016) are situated in the North Sea with strong connectivity to some of North Americas largest shipping centres *e.g.* New York-New Jersey or the Port of Virginia (Seebens, Gastner and Blasius, 2013). If *M. leidyi* was able to invade through ballast water transfer and avoid bottlenecking via high propagule pressure and/or continuous gene flow, other marine invaders may experience similarly facilitated establishment.

4.2 Diversity and Demography

Earlier theory of invasion biology proposed invasive populations to heavily suffer from reduced genetic diversity due to founder effects (Baker and Stebbins, 1965; Allendorf and Lundquist, 2003). However, a variety of evidently successful invasive species both terrestrial and marine have led to the formulation of the genetic paradox of invasions (Sax and Brown, 2000) that describes the contradictory establishment of non-indigenous populations with low genetic diversity. Recently, the paradox itself has been challenged (Roman and Darling, 2007; Estoup *et al.*, 2016), with several reports of invasive populations that did not or currently do not suffer from reduced genetic diversity with regards to their source populations (Kolbe *et al.*, 2004). Especially for the case of marine invasive species the majority of studies (63%) report no significant reduction in genetic diversity (Roman and Darling, 2007). Here we present the first comparison of genetic diversity between native and invasive marine populations at the whole-genome scale, demonstrating that all scenarios; conservation, reduction and to some extent increase of genomic diversity may occur.

4.2.1 Diversity Conservation

Indigenous Miami samples and their invasive counterparts from the Black Sea have demonstrated remarkable similarity across all indicators of genomic diversity. From levels of heterozygosity and inbreeding to the number of segregating sites, Miami and Varna samples have shown no significant difference in population means. Only for genome-wide nucleotide diversity did Miami exhibit a significantly higher mean though the absolute difference is minimal. More astonishing than equality of means is perhaps the conservation of the full spectrum of genomic diversity. For both Watterson's theta and Nei's nucleotide diversity the distributions across the genome are close to identical demonstrating that low- as well as high-diversity regions are maintained in the invasive population. These results are in contrast to those found by previous *M. leidyi* studies, which reported a significant decrease in genetic diversity in the Black Sea (Reusch *et al.*, 2010; Bolte *et al.*, 2013; Ghabooli *et al.*, 2013; Bayha *et al.*, 2015). However, it should be noted that these studies relied on comparatively limited markers and that individuals were sampled 7-10 years prior to those sequenced here. Highly variable regions such as internally transcribed spacer (ITS) sequences, the mitochondrial cytochrome *b* gene or microsatellites, which were all used in prior studies, are only representative of the upper bounds of the full spectrum of genomic diversity. Such markers may therefore overestimate the true mean or differences between true means, giving an incomplete representation of genome-wide diversity. In fact, previously reported nucleotide diversities are larger by approximately a factor 2 when compared to the genomic means reported here.

An initially bottlenecked Black Sea population, which regenerated genomic diversity during the past decade either through recurrent or multiple invasion waves, would explain the diversity equivalence between Varna and Miami. However, an inbreeding coefficient of zero speaks against identity by descent that would result from a small founder population. In addition, a significant pairwise F_{ST} is contrary to substantial gene flow between Florida and the Black Sea whilst close matching of heterozygosity, and nucleotide diversity

spectra make admixture of a secondary, unsampled source region unlikely. The results presented here are most in favour of a large invasion event with high propagule pressure that transferred the full spectrum of genomic variance to the invasive population that has since diverged allopatrically via genetic drift. This interpretation can be reconciled with the theory of *M. leidyi* invading the Black Sea in the ballast water of Soviet vessels during the 1980s. Heavy shipping traffic between Cuba and the U.S.S.R.'s major Black Sea ports of Sevastopol, Feodosia and Novorossiysk (Miller, 1989), may have led to the exchange of vast quantities of ballast water. Incidentally, all three ports are within 200km of the Bay of Sudak where *M. leidyi* was first sighted (Pereladov, 1988). However, shipping traffic found an abrupt end in 1990, just before the collapse of the Soviet Union, which would also explain the current lack of gene flow and continued divergence.

Besides pure diversity indicators, genome-wide distributions of Tajima's *D* are also near identical for Miami and Varna samples. While non-zero values of *D* may be indicative of directional or balancing selection in specific regions, the genome-wide distribution is more suggestive of demography as changes in population size affect neutral regions also. Here, the negative means displayed by both Miami and Varna imply a subtle overrepresentation of low-frequency alleles, also observable in their respective AFS. This excess of rare relative to common SNPs is suggestive of a recent population expansion where even low-frequency polymorphisms are not lost to drift. Furthermore, Miami and Varna exhibit equally low LD implying weak haplotype structure. Whilst long-rang LD was likely broken down via recombination, founder effects would have led Varna to retain high short-range LD. This is not the case and instead, Varna shows no significant indication of conceivable founder effects or current lack of diversity compared to Miami. In fact, Varna displays a contemporary effective population size of the same order as Miami with both populations expanding during recent times, consistent with the other demographic indicators.

4.2.2 Diversity Reduction

As much as the primary invasion into the Black Sea defies classical invasion genetic theory, the apparent range expansion into the Mediterranean seems to confirm its hypotheses. Regardless of whether the Gulf of Mexico or the Black Sea seeded the Western Mediterranean, Villefranche specimens exhibited considerably dissimilar diversity indicators to both source regions. Genome-wide Watterson's theta and Nei's nucleotide diversity were significantly lower, indicating fewer segregating sites and pairwise nucleotide differences between Villefranche individuals, even after correction for sample size. Both mean individual and population-wide heterozygosities were higher than those of Miami and Varna, which could be interpreted as a diversity increase but is most likely due to a lack of rare polymorphisms. Such absence of low-frequency SNPs, visible in the AFS also, leads to an increase in the proportion of heterozygous loci, thus inflating heterozygosity indices. In combination, a low number of segregating sites and high heterozygosity are indicative of strong haplotype structure. Under the assumption that Villefranche was seeded from the Black Sea, these results therefore point towards a bottleneck or selective sweep having restored haplotype structure by removing rare variants.

Finding a lower number of rare, private polymorphisms might be an artefact attributable to lower sequencing depth though other, more robust, indicators also point towards a bottleneck or sweep. A significant, positive genome-wide Tajima's D, for instance, suggests a recent population contraction. Since D measures discrepancy between θ_W and π, it is a relative index and independent of the absolute number of polymorphisms. Strong haplotype structure can also be observed in the plot of LD decay where Villefranche reveals higher correlation coefficients than Miami and Varna, both at short and long physical distances. Resulting estimates N_e confirm Villefranche to have less than a third of the size of Miami and Varna populations. Finally, the marked similarity between the LD decay curve shapes of Varna and Villefranche is further evidence in favour of a range expansion from the Black Sea with population contraction leading to an upshift in the Villefranche curve without distortion.

Two prior studies have also demonstrated the decrease of genetic diversity, arguing for a stepping-stone invasion dynamic from the Gulf of Mexico to the Black Sea and subsequently into the Mediterranean (Bolte *et al.*, 2013; Bayha *et al.*, 2015). Whereas a noticeable decrease in genomic diversity during the first "step" is not evident, Villefranche, at the limits of the Mediterranean expansion front, does show a substantial signal at the genome-wide scale. This allows for several interpretations. Firstly, as *M. leidyi* migrated westwards successive founder events may have kept effective population size low and eroded rare polymorphisms through drift. Secondly, with no evidence of allelic recycling, *M. leidyi* individuals acclimatized to the Black Sea would have had to re-adapt to the Mediterranean. Contrasting environmental conditions, such as extreme salinities (Stanev, 1990; Poulos, Drakopoulos and Collins, 1997), may have triggered a selective sweep, eradicating most genomic diversity in the process. Finally, even in the absence of selection or multiple founder events, costly acclimatization via phenotypic plasticity may be responsible for extended lag phases in a founding population. Such periods of low population size prior to expansion have been shown to significantly reduce genetic diversity in marine invasive populations through drift (Gaither, Toonen and Bowen, 2012). In fact, a time lag of over 15 years between the first sighting (Kideys and Niermann, 1994) and establishment of *M. leidyi* in the Mediterranean (Boero *et al.*, 2009; Fuentes *et al.*, 2009; Galil, Kress and Shiganova, 2009) would support this hypothesis. Whether genetic drift in small stepping-stone populations, a strong selective sweep, an extended lag phase or a combination of scenarios is the reason for reduced genomic diversity in Western Mediterranean, remains to be elucidated.

4.2.3 Diversity Increase and Demographic Influence

Perhaps the most interesting pattern of native-invasive genomic diversity is displayed by the northern invasion system. Although closely related, invasive Sylt individuals reveal slightly higher genomic diversity than their native counterparts. PCA results of northern samples in isolation show Sylt specimens clustered broadly whilst Woods Hole displayed low variance, with the second PC

almost exclusively representing variance in the Sylt population. Both Watterson's theta and nucleotide diversity show genome-wide distributions that are skewed towards larger values in Sylt whilst Woods Hole displays a distinct excess of low diversity regions. These findings are in contrast to previous investigations that found a significant reduction in genetic diversity in the northern invasive range (Reusch *et al.*, 2010; Bolte *et al.*, 2013; Ghabooli *et al.*, 2013). Still, the most recent study by Bayha *et al.* (2015) did find equivalent genetic diversity with North Sea samples showing higher, albeit insignificant indices.

Increased genomic diversity in the invasive range is often accounted for by multiple introductions from distinct source regions (Kolbe *et al.*, 2004; Darling *et al.*, 2008; Facon *et al.*, 2008). However, a particularly low pairwise F_{ST} and low eigenvalues suggest source matching is adequate and admixture of further, isolated populations unlikely. An alternative explanation for the slight discrepancy between Woods Hole and Sylt could be diversity reduction in the native range rather than an invasion specific increase in the North Sea. Tight clustering of Woods Hole individuals and multiple conserved genomic regions could be indicative of purifying selection and strong local adaptation. However, negative inbreeding coefficients, such as those observed for Woods Hole, are uncharacteristic for highly adapted populations that often suffer from inbreeding depression (Verhoeven *et al.*, 2011).

Costello *et al.* (2006) observed the population dynamics of *M. leidyi* in Narragansett Bay, RI and demonstrated that whilst *M. leidyi* is predominantly absent during the winter months, small bays serve as isolated overwintering sites for a minor number of individuals. These localised refugia subsequently re-seed open-bay waters during spring giving rise to a large-scale bloom deriving from only a few individuals. The authors propose this seasonal decline and re-seeding to take place in most temperate regions occupied by *M. leidyi*. Indeed, annual fluctuations in population size have been confirmed several times in the invasive range (Faasse and Bayha, 2006; Fuentes *et al.*, 2009; Riisgård *et al.*, 2012) and it is likely that both the Woods Hole and Sylt populations sampled

here are also subject to such demography. Estimates of effective population size for Woods Hole samples are the most erratic. Whilst a regular period cannot be identified, fluctuations are evident with a prominently steep decline in the recent past. A positive genome-wide Tajima's *D* and an absence of rare SNPs also suggest the Woods Hole population to have undergone a recent bottleneck. In addition, a large variance in *D* is indicative of a heterogeneous genome possibly shaped by irregular periods of population expansion and contraction. A similar pattern can be seen in Sylt, which might be equally affected by regular fluctuations in population size though estimates of N_e do not agree. Whilst we cannot make definitive claims, it is plausible that seasonal demography is primarily responsible for shaping local genomic diversity in the northern populations, more so than invasion dynamics. In fact, it has been suggested that *M. leidyi*'s ability to rebound from low densities via source-sink dynamics is the principal reason for its invasive success (Costello *et al.*, 2012). Varying severity of these demographic fluctuations alongside evidence for a recent population contraction in Woods Hole might explain the discrepancy in genomic diversity observed between northern populations, as well as between previous studies. Demography of non-indigenous populations is often addressed when discussing similar results in the field of invasion biology, however short-term native range dynamics are oftentimes ignored and comparisons may therefore be biased. The outcome above highlights the importance of focussing on both native and invasive range demography in order to draw informed conclusions.

4.3 Contrasting Independent Invasions

4.3.1 Genomic Parallelism and Deviation

Evident from the results presented here, the two *M. leidyi* invasion systems exhibit distinct patterns, likely shaped by unique demographic processes in the native range, seasonality, local environmental conditions, time since establishment and stepping-stone dynamics. Hence, few parallels can be drawn between the northern and southern invasion events. The attempt to detect molecular parallelism in the two invasion systems via the projection of southern populations onto the principal components derived from northern individuals did not yield obvious results. Minimal separation of the southern specimens might suggest there to be a set of SNPs divergent between native and invasive populations yet common to both invasion systems. However, outlier analyses or a discriminant analysis of principal components with pre-defined groups might yield more conclusive results.

While the focus of this study was to contrast invasive populations with their native counterparts, the most prominent differences were observed between northern and southern populations. Given their deep genomic divergence, a dominant signal was not surprising. Both Miami and Varna displayed considerably different patterns for almost all indicators when compared to Woods Hole and Sylt. Significantly higher genomic diversity, lower heterozygosity and a steeper curve in the AFS suggest southern populations contain more low-frequency polymorphisms than northern individuals. These results are encouraging with regard to mapping bias. Samples that are genetically more distant from the reference genome, like the southern specimens, tend to exhibit lower mapping rates due to unsuccessful mapping in highly diverged regions (Brandt *et al.*, 2015). This, in turn leads to lower sequencing depth, reduced call rates and consequently less power to detect rare variants. The fact that Miami and Varna display a higher proportion of rare polymorphisms despite reduced mapping rates suggests that the biological signal overrides any bias, and the difference to the northern populations is at most underestimated in this study.

Assuming a genuine signal, the question remains as to how such a pattern has formed. Seasonal fluctuations of population size in temperate regions (Costello *et al.*, 2006) might be responsible for purging novel, low-frequency SNPs in the north. Genetic drift during periods of low abundance would then skew the AFS towards higher allele frequencies. Another possible explanation is that Miami is the most ancestral of the populations sampled. Such populations typically exhibit higher diversity with a large proportion of rare polymorphisms due to the accumulation of multiple mutations and the lack of historic bottlenecks (Yu *et al.*, 2002). In fact, Miami also presents very low short-range LD, which necessarily required numerous recombination and/or mutation events to erode. In line with this argument lower genomic diversity and higher LD in Woods Hole may possibly be due to bottlenecking during a range expansion into higher latitudes after, for instance, the last glacial maximum. Up until 18,000 BCE the Laurentide ice sheet covered coastal waters of the Northwest Atlantic down to a latitude of approximately 40° N, roughly the height of New York, NY (Dyke and Prest, 1987). Colonisation of these northern habitats might have taken place shortly thereafter, possibly aided by fluctuations in Gulf Stream velocity (Kaneps, 1979), which may have allowed planktonic organisms to cross the oceanographic barrier at Cape Hatteras, NC. However, coalescence analyses will be required to confirm these claims and date the nature and historic time point of divergence between the northern and southern haplogroups more accurately.

4.3.2 The Simultaneous Hermaphrodite Advantage

The reproductive mode of *M. leidyi* has often been cited as a fundamental attribute, determinant in its invasion success (Costello *et al.*, 2012; Bayha and Graham, 2014). Understandably, a facultative simultaneous hermaphrodite can reap the benefits of sexual reproduction when possible yet re-establish itself from extremely low abundances through selfing. Both characteristics seem particularly appealing to a marine invader that must found a colony from small inoculates whilst also having to adapt to novel environmental conditions.

Using a whole-genome dataset sourced from two independent invasion systems we found no evidence of *M. leidyi* having undertaken substantial selfing, neither in the invasive nor in the native range. A minimal number of SNPs showed departure from HWE whilst inbreeding coefficients were insignificantly different from zero for all populations. In fact, Woods Hole and Sylt even displayed marginally negative F_{IS} suggestive of some degree of non-random outbreeding. Although low F_{IS} values are atypical for a planktonic marine invertebrate (Addison and Hart, 2005) it appears that self-fertilisation is not employed by *M. leidyi* during the invasion process. Recent work on the reproductive mode of *M. leidyi* has shown that selfing is costly in the ctenophore. Embryos of self-fertilised individuals displayed reduced viability compared to those that spawned in pairs (Sasson and Ryan, 2016). If propagule pressure was sufficiently high during first establishment, which seems to be the case for the Black and North Sea invasions, costly self-fertilisation might not have been necessary resulting in outbred populations as evidenced here.

Zero inbreeding coefficients and populations in HWE stand in contrast with previous studies where significant heterozygote deficit was shown and argued to be the result of precisely self-fertilisation during invasion (Reusch *et al.*, 2010; Bolte *et al.*, 2013; Ghabooli *et al.*, 2013; Bayha *et al.*, 2015). Discrepancies may however be due to the blooming nature of *M. leidyi* and the associated sampling error. Rapid population expansion, especially in secluded bays or harbours with little advection (Purcell *et al.*, 2001), can generate localised genetic substructure, even in a planktonic organism (Bayha *et al.*, 2015). Sampling during these blooms may consequently lead to inflated inbreeding coefficients and heterozygote deficits due to the Wahlund effect. Here, insignificant F_{IS} and populations in HWE give no indication of self-fertilisation or local sub-structure during the time of sampling.

5. Conclusion

By applying a high-density SNP set the population structure, associated diversity patterns and demography of the *Mnemiopsis leidyi* invasion system were elucidated. To the best of our knowledge this is the first study comparing native and invasive populations of a marine invader at the whole-genome scale. Employing a large number of markers allowed for accurate and highly resolved estimates of population differentiation, the evaluation of precise genome-wide diversity spectra and a more robust inference of population-specific demography.

Genetic structure proved to be hierarchical, suggestive of distinct invasion profiles and largely corresponding to time since establishment. We confirm the existence of two independent invasion events into southern and northern European waters, distinguishable by high divergence in the native range. Source regions were confirmed to be in the vicinity of the Floridian Peninsula and the coast of New England though more extensive sampling would be required to pinpoint exact locations. Our results advocate that *M. leidyi* first arrived in the Black Sea where it fully established and has since allopatrically diverged from its source population. Subsequent range expansion into and throughout the Mediterranean Basin took place more recently, accompanied by further divergence away from both Black Sea and the native population. Secondary, trans-Atlantic invasion events and ensuing admixture of native individuals does not seem to occur in the Western Mediterranean. Invasion of the North Sea appears to have taken place most recently with minimal divergence between native and invasive populations. Such similarity is suggestive of large inoculates and/or sustained gene flow between regions, two bothersome interpretations that warrant further investigation of current transport vectors.

Prominent hypotheses regarding diversity patterns between native and invasive marine populations were tested for the first time using whole-genome data. We present evidence of genome-wide diversity increase, reduction as well as conservation relative to source populations, offering explanations in accordance with demographic indicators. Western Mediterranean samples supported the classic paradigm of diversity reduction, likely caused by successive founder events during a range expansion, albeit successful adaptation following a selective sweep is also plausible. The Black Sea population exhibited remarkable equivalence of genomic diversity in favour of a modern view of invasion genetics. Introduction from multiple source regions is a possibility though close matching of the genome-wide spectrum of diversity points to high propagule pressure during initial establishment having transferred the full breadth of diversity. North Sea individuals displayed moderately higher genomic diversity than their native counterparts, further challenging the paradigm that young founder populations are genetically deprived. Whilst incorrect source matching might be the cause, demographic and ecological indicators point to seasonal fluctuations in population size, rather than invasive processes, that shape diversity in temperate regions.

Opposing patterns of diversity at the whole-genome scale, yet within a single species, highlight the importance of particular invasion dynamics and specific context. Hence, the question remains whether genomic diversity has an impact on invasion success or whether it is just a correlate of more significant, extrinsic factors. Density-dependent ecological processes, propagule pressure and expansion lag times may be much more influential and shape genomic diversity in parallel. Especially highly plastic species such as *M. leidyi* may not be affected by its intrinsic diversity, coping with novel environmental conditions through acclimatisation rather than adaptation. In such organisms one would expect elevated levels of cloning or selfing and the forfeiting of the benefits of sexual reproduction in favour of increased fecundity. Nonetheless, the simultaneous hermaphrodite *M. leidyi* shows no signal of inbreeding or Hardy-Weinberg disequilibrium across the genome, even suggesting active maintenance of diversity in Woods Hole.

It is clear further work is required to disentangle the confounding factors of demography, ecology and genetics in determining the success of marine invaders. With its well-defined population structure, spatially and temporally independent invasion events and genomic resources, the invasive ctenophore *Mnemiopsis leidyi* can be harnessed as a valuable model system. Through the elaboration of whole-genome re-sequencing data several exciting aspects in the field of invasion genomics may now be addressed in the near future. For instance, the relative contributions of genotype matching and novel mutations in local adaptation to the invasive range may be elucidated. Detection of molecular parallelism through outlier analyses or discriminant analyses of principal components (DAPC) may lead to the discovery of "invasive genes" that are independently selected for in separate invasion events. Finally, a coalescence approach using the entire genome-wide data set will yield a clearer picture of demography through time, possibly revealing extant gene-flow and resolving questions of ancestry and historic admixture.

Acknowledgements

I would like to give special thanks to my supervisor Conny Jaspers for getting me interested in the *"Mnops"* and sticking with it even when things didn't work out the first, second or even third time. Thank you for entrusting an NGS project to me even though I had little previous experience and thank you for making it possible that I could sample along the Cotê d'Azur, the "Golden Sands" of the Black Sea and on the island of Sylt. Thank you also for preparing and receiving samples in the US, for being open to my requests and all of the extra work that didn't even make it into this thesis.

A thank you to Thorsten Reusch for giving the necessary input when it counted and keeping me focussed on the main questions.

Thank you Sven Künzel for supervising the sequencing process and the DNA exchanges at the gas station. Thank you for countless phone calls and for how quickly you managed to get everything done, despite needing several repeats and having to tackle many "sequencing blockades".

Thank you Till Bayer for countless opinions and tips with everything bioinformatics related, thanks for joining all of the meetings and giving me "primers" on all techniques.

Many thanks go to Morten Limborg for the brief but extremely helpful meetings and obviously for continuing what I have started.

I would also like to thank Bill Browne from the University of Miami for supplying me with *M. leidyi* samples even at the shortest of notices only to then have to collect more and send them himself a few months later. Thank you also for the tips on DNA extraction and the uncompromisingly friendly exchange.

Thank you Kremena Stefanova from the Institute of Oceanology-BAS in Varna for the warm welcome in Bulgaria and the time taken to help me sample. Thank you for lunches at the beach and the preparation of dozens of specimens.

The same thanks go to Fabien Lombard and Simon Ramondenc from the Laboratoire d'Océanologie de Villefranche-sur-Mer for sampling and storing *Mnemiopsis* for me to collect, they really made a big difference.

Many thanks go to Birgit Hussel at the AWI Sylt for sending multiple samples and showing me around the institute when I had to go myself.

Obviously many, many thanks to my office mates and friends; Lara for passively motivating me by being super-focused, Jenny for putting everybody in a good mood on the rainiest of days, Vero for always answering the phone when nobody else does, Verena for the much-needed distractions from work, Kosmas for helpful discussion and not-so-helpful jokes and Serra for the great company during many late office nights.

Thank you so much Michu for knowing exactly when to calm me down, from nearby or even from thousands of kilometers away.

Finally, an enormous thank you to my parents without whom I probably would not have sought or been able to get to this point.

Funding

This M.Sc. thesis received support from the MOBILEX grant (DFF-1325-00102B) by which Cornelia Jaspers was financed through the European Union's Seventh Framework Program for research, technological development and demonstration - Marie Curie and the Danish Council for Independent Research. Additional resources were provided through the grant (CP1539, PI C. Jaspers) from the Cluster of Excellence 80 'The Future Ocean'. 'The Future Ocean' is funded through the framework of the Excellence Initiative by the Deutsche Forschungsgemeinschaft (DFG) on behalf of the German federal and state governments.

Bibliography

Addison, J. A. and Hart, M. W. (2005) 'Spawning, copulation and inbreeding coefficients in marine invertebrates', *Biology Letters*, 1(4), pp. 450–453. doi: 10.1098/rsbl.2005.0353.

Allee, W. C. (1931) *Animal aggregations, a study in general sociology.*, *University of Chicago Press, Chicago*. Chicago: The University of Chicago Press,. doi: 10.5962/bhl.title.7313.

Allendorf, F. W. and Lundquist, L. L. (2003) 'Introduction: Population biology, evolution, and control of Invasive Species', *Conservation Biology*. Blackwell Science, Inc., pp. 24–30. doi: 10.1046/j.1523-1739.2003.02365.x.

Arnold, M. L., Ballerini, E. S. and Brothers, A. N. (2012) 'Hybrid fitness, adaptation and evolutionary diversification: lessons learned from Louisiana Irises', *Heredity*, 108(3), pp. 159–166. doi: 10.1038/hdy.2011.65.

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S. and DePristo, M. A. (2013) 'From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline', in *Current Protocols in Bioinformatics*. Hoboken, NJ, USA: John Wiley & Sons, Inc., p. 11.10.1-11.10.33. doi: 10.1002/0471250953.bi1110s43.

Bailey, S. F., Rodrigue, N. and Kassen, R. (2015) 'The effect of selection environment on the probability of parallel evolution', *Molecular Biology and Evolution*. Oxford University Press, 32(6), pp. 1436–1448. doi: 10.1093/molbev/msv033.

Baker, H. G. and Stebbins, G. L. (1965) 'The Genetics of Colonizing Species', *1965*, pp. 147–168.

Baker, L. D. and Reeve, M. R. (1974) 'Laboratory culture of the lobate ctenophore Mnemiopsis mccradyi with notes on feeding and fecundity', *Marine Biology*, 26(1), pp. 57–62. doi: 10.1007/BF00389086.

Bass, A. J., Dabney, A. and Robinson, D. D. (2015) 'qvalue: Q-value estimation for false discovery rate control.', *R package version 2.8.0*.

Bax, N., Williamson, A., Aguero, M., Gonzalez, E. and Geeves, W. (2003) 'Marine invasive alien species: a threat to global biodiversity', *Marine Policy*. Elsevier, 27(4), pp. 313–323. doi: 10.1016/S0308-597X(03)00041-1.

Bayha, K. M., Chang, M. H., Mariani, C. L., Richardson, J. L., Edwards, D. L., DeBoer, T. S., Moseley, C., Aksoy, E., Decker, M. B., Gaffney, P. M., Harbison, G. R., McDonald, J. H. and Caccone, A. (2015) 'Worldwide phylogeography of the invasive ctenophore Mnemiopsis leidyi (Ctenophora) based on nuclear and mitochondrial DNA data', *Biological Invasions*, 17(3), pp. 827–850. doi: 10.1007/s10530-014-0770-6.

Bayha, K. M. and Graham, W. M. (2014) 'Nonindigenous Marine Jellyfish: Invasiveness, Invasibility, and Impacts', in *Jellyfish Blooms*. Dordrecht: Springer Netherlands, pp. 45–77. doi: 10.1007/978-94-007-7015-7_3.

Bell, M. A. and Aguirre, W. E. (2013) 'Contemporary evolution, allelic recycling, and adaptive radiation of the threespine stickleback', *Evolutionary Ecology Research*. Evolutionary Ecology, Ltd., 15(4), pp. 377–411.

Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R. and Bernatchez, L. (2015) 'RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (Homarus americanus)', *Molecular Ecology*, 24(13), pp. 3299–3315. doi: 10.1111/mec.13245.

Bernardi, G., Azzurro, E., Golani, D. and Miller, M. R. (2016) 'Genomic signatures of rapid adaptive evolution in the bluespotted cornetfish, a Mediterranean Lessepsian invader', *Molecular Ecology*, 25(14), pp. 3384–3396. doi: 10.1111/mec.13682.

Besiktepe, S. T., Sur, H. I. do., Özsoy, E., Latif, M. A., Oguz, T. and Ünlüata, Ü. (1994) 'The circulation and hydrography of the Marmara Sea', *Progress in Oceanography*, 34(4), pp. 285–334. doi: 10.1016/0079-6611(94)90018-3.

Bhatia, G., Patterson, N., Sankararaman, S. and Price, a. L. (2013) 'Estimating and interpreting F_{ST}: The impact of rare variants', *Genome Research*, 23(9), pp. 1514–1521. doi: 10.1101/gr.154831.113.

Bilio, M. and Niermann, U. (2004) 'Is the comb jelly really to blame for it all? Mnemiopsis leidyi and the ecological concerns about the Caspian Sea', *Marine Ecology Progress Series*, 269, pp. 173–183. doi: 10.3354/meps269173.

Blakeslee, A. M. H., McKenzie, C. H., Darling, J. A., Byers, J. E., Pringle, J. M. and Roman, J. (2010) 'A hitchhiker's guide to the Maritimes: anthropogenic transport facilitates long-distance dispersal of an invasive marine crab to Newfoundland', *Diversity and Distributions*. Blackwell Publishing Ltd, 16(6), pp. 879–891. doi: 10.1111/j.1472-4642.2010.00703.x.

Bock, D. G., Caseys, C., Cousens, R. D., Hahn, M. A., Heredia, S. M., Hübner, S., Turner, K. G., Whitney, K. D. and Rieseberg, L. H. (2015) 'What we still don't know about invasion genetics.', *Molecular ecology*, 24(9), pp. 2277–97. doi: 10.1111/mec.13032.

Boero, F., Putti, M., Trainito, E., Prontera, E., Piraino, S. and Shiganova, T. (2009) 'First records of Mnemiopsis leidyi (Ctenophora) from the Ligurian, Thyrrhenian and Ionian Seas (Western Mediterranean) and first record of Phyllorhiza punctata (Cnidaria) from the Western Mediterranean', *Aquatic Invasions*, 4(4), pp. 675–680. doi: 10.3391/ai.2009.4.4.13.

Bollback, J. P. and Huelsenbeck, J. P. (2009) 'Parallel genetic evolution within and between bacteriophage species of varying degrees of divergence', *Genetics*, 181(1), pp. 225–234. doi: 10.1534/genetics.107.085225.

Bolte, S., Fuentes, V., Haslob, H., Huwer, B., Thibault-Botha, D., Angel, D., Galil, B., Javidpour, J., Moss, A. and Reusch, T. (2013) 'Population genetics of the invasive ctenophore Mnemiopsis leidyi in Europe reveal source–sink dynamics and secondary dispersal to the Mediterranean Sea', *Marine Ecology Progress Series*, 485, pp. 25–36. doi: 10.3354/meps10321.

Bortolotto, E., Bucklin, A., Mezzavilla, M., Zane, L. and Patarnello, T. (2011) 'Gone with the currents: lack of genetic differentiation at the circum-continental scale in the Antarctic krill Euphausia superba.', *BMC genetics*, 12(1), p. 32. doi: 10.1186/1471-2156-12-32.

Bradic, M., Teotónio, H., Borowsky, R. L., Teotonio, H. and Borowsky, R. L. (2013) 'The population genomics of repeated evolution in the blind cavefish astyanax mexicanus', *Molecular Biology and Evolution*. Oxford University Press, 30(11), pp. 2383–2400. doi: 10.1093/molbev/mst136.

Brandt, D. Y. C., Aguiar, V. R. C., Bitarello, B. D., Nunes, K., Goudet, J. and Meyer, D. (2015) 'Mapping Bias Overestimates Reference Allele Frequencies at the HLA Genes in the 1000 Genomes Project Phase I Data.', *G3 (Bethesda, Md.)*, 5(5), pp. 931–41. doi: 10.1534/g3.114.015784.

Briski, E., Bailey, S. A., Casas-Monroy, O., DiBacco, C., Kaczmarska, I., Lawrence, J. E., Leichsenring, J., Levings, C., MacGillivary, M. L., McKindsey, C. W., Nasmith, L. E., Parenteau, M., Piercey, G. E., Rivkin, R. B., Rochon, A., Roy, S., Simard, N., Sun, B., Way, C., Weise, A. M. and MacIsaac, H. J. (2013) 'Taxon- and vector-specific variation in species richness and abundance during the transport stage of biological invasions', *Limnology and Oceanography*, 58(4), pp. 1361–1372. doi: 10.4319/lo.2013.58.4.1361.

Burton, O. J. and Travis, J. M. J. (2008) 'The Frequency of Fitness Peak Shifts Is Increased at Expanding Range

Margins Due to Mutation Surfing', *Genetics*, 179(2), pp. 941–950. doi: 10.1534/genetics.108.087890.

Carlton, J. T. and Geller, J. B. (1993) 'Ecological Roulette: The Global Transport of Nonindigenous Marine Organisms', *Science*, 261(5117), pp. 78–82. doi: 10.1126/science.261.5117.78.

Cattell, R. B. (1966) 'The scree test for the numbers of factors', *Multivariate Behavioral Research*, 1(2), pp. 245–276. doi: 10.1207/s15327906mbr0102.

Chown, S. L., Hodgins, K. A., Griffin, P. C., Oakeshott, J. G., Byrne, M. and Hoffmann, A. A. (2015) 'Biological invasions, climate change and genomics', *Evolutionary Applications*. Wiley-Blackwell, 8(1), pp. 23–46. doi: 10.1111/eva.12234.

Colautti, R. I. and Lau, J. A. (2015) 'Contemporary evolution during invasion: evidence for differentiation, natural selection, and local adaptation.', *Molecular ecology*, 24(9), pp. 1999–2017. doi: 10.1111/mec.13162.

Conte, G. L., Arnegard, M. E., Peichel, C. L. and Schluter, D. (2012) 'The probability of genetic parallelism and convergence in natural populations.', *Proceedings. Biological sciences / The Royal Society*, 279(1749), pp. 5039–47. doi: 10.1098/rspb.2012.2146.

Corbin, L. J., Liu, A. Y. H., Bishop, S. C. and Woolliams, J. A. (2012) 'Estimation of historical effective population size using linkage disequilibria with marker data', *Journal of Animal Breeding and Genetics*, 129(4), pp. 257–270. doi: 10.1111/j.1439-0388.2012.01003.x.

Costello, J. H., Bayha, K. M., Mianzan, H. W., Shiganova, T. A. and Purcell, J. E. (2012) 'Transitions of Mnemiopsis leidyi (Ctenophora: Lobata) from a native to an exotic species: a review', *Hydrobiologia*, 690(1), pp. 21–46. doi: 10.1007/s10750-012-1037-9.

Costello, J. H., Sullivan, B. K., Gifford, D. J., Keuren, D. Van and Sullivan, L. J. (2006) 'Seasonal refugia, shoreward thermal amplification, and metapopulation dynamics of the ctenophore Mnemiopsis leidyi in Narragansett Bay, Rhode Island', *Limnology and Oceanography*, 51(4), pp. 1819–1831. doi: 10.4319/lo.2006.51.4.1819.

Crooks, J. A. (2002) 'Characterizing ecosystem-level consequences of biological invasions: the role of ecosystem engineers', *Oikos*. Munksgaard International Publishers, 97(2), pp. 153–166. doi: 10.1034/j.1600-0706.2002.970201.x.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G. and Durbin, R. (2011) 'The variant call format and VCFtools', *Bioinformatics*. Oxford University Press, 27(15), pp. 2156–2158. doi: 10.1093/bioinformatics/btr330.

Darling, J. A., Bagley, M. J., Roman, J., Tepolt, C. K. and Geller, J. B. (2008) 'Genetic patterns across multiple introductions of the globally invasive crab genus Carcinus.', *Molecular ecology*, 17(23), pp. 4992–5007. doi: 10.1111/j.1365-294X.2008.03978.x.

Díez-Del-Molino, D. (2016) 'Genetic diversity and population structure of the non-native Eastern mosquitofish (Gambusia holbrooki) in Mediterranean streams', (August). doi: 10.13140/RG.2.1.1499.3128.

Dlugosch, K. M., Anderson, S. R., Braasch, J., Cang, F. A. and Gillette, H. D. (2015) 'The devil is in the details: genetic variation in introduced populations and its contributions to invasion.', *Molecular ecology*, 24(9), pp. 2095–111. doi: 10.1111/mec.13183.

Dlugosch, K. M. and Parker, I. M. (2008) 'Founding events in species invasions: Genetic variation, adaptive evolution, and the role of multiple introductions', *Molecular Ecology*, 17(1), pp. 431–449. doi: 10.1111/j.1365-294X.2007.03538.x.

Drake, J. M. (2006) 'Heterosis, the catapult effect and establishment success of a colonizing bird', *Biology Letters*, 2(2), pp. 304–307. doi: 10.1098/rsbl.2006.0459.

Dyke, A. S. and Prest, V. K. (1987) 'Late Wisconsinan and Holocene History of the Laurentide Ice Sheet', *Géographie physique et Quaternaire*, 41(2), p. 237. doi: 10.7202/032681ar.

Estoup, A., Ravigné, V., Hufbauer, R., Vitalis, R., Gautier, M. and Facon, B. (2016) 'Is There a Genetic Paradox of Biological Invasion?', *Annual Review of Ecology, Evolution, and Systematics*, 47(1), pp. 51–72. doi: 10.1146/annurev-ecolsys-121415-032116.

Eurostat (2016) 'Energy, transport and environment indicators, 2016 edition'. doi: 10.2785/260003.

Excoffier, L., Foll, M. and Petit, R. J. (2009) 'Genetic Consequences of Range Expansions', *Annual Review of Ecology, Evolution, and Systematics*. Annual Reviews, 40(1), pp. 481–501. doi: 10.1146/annurev.ecolsys.39.110707.173414.

Faasse, M. A. and Bayha, K. M. (2006) 'The ctenophore Mnemiopsis leidyi A. Agassiz 1865 in coastal waters of the Netherlands: an unrecognized invasion?', *Aquatic Invasions*, 1(4), pp. 270–277. doi: 10.3391/ai.2006.1.4.10.

Facon, B., Pointier, J.-P., Jarne, P., Sarda, V. and David, P. (2008) 'High genetic variance in life-history strategies within invasive populations by way of multiple introductions.', *Current biology : CB*, 18(5), pp. 363–7. doi: 10.1016/j.cub.2008.01.063.

Forsman, A. (2014) 'Effects of genotypic and phenotypic variation on establishment are important for conservation, invasion, and infection biology.', *Pnas*, 111(1), pp. 302–7. doi: 10.1073/pnas.1317745111.

Fuentes, V. L., Atienza, D., Gili, J. M. and Purcell, J. E. (2009) 'First records of mnemiopsis leidyi A. Agassiz 1865 off the NW mediterranean coast of Spain', *Aquatic Invasions*, 4(4), pp. 671–674. doi: 10.3391/ai.2009.4.4.12.

Funk, W. C., McKay, J. K., Hohenlohe, P. A. and Allendorf, F. W. (2012) 'Harnessing genomics for delineating conservation units', *Trends in Ecology and Evolution*, pp. 489–496. doi: 10.1016/j.tree.2012.05.012.

Gaither, M. R., Toonen, R. J. and Bowen, B. W. (2012) 'Coming out of the starting blocks: extended lag time rearranges genetic diversity in introduced marine fishes of Hawai'i', *Proceedings of the Royal Society B: Biological Sciences*, 279(1744), pp. 3948–3957. doi: 10.1098/rspb.2012.1481.

Galil, B. S., Kress, N. and Shiganova, T. A. (2009) 'First record of Mnemiopsis leidyi A. Agassiz, 1865 (Ctenophora; Lobata; Mnemiidae) off the Mediterranean coast of Israel', *Aquatic Invasions*, 4(2), pp. 357–360. doi: 10.3391/ai.2009.4.2.8.

Gascuel, O. (1996) 'BIONJ: An Improved Version of the NJ Algorithm Based on a Simple Model of Sequence Data', *Molecular Biology and Evolution*, 14(7), pp. 685–695.

GESAMP (1997) *Opportunistic settlers and the problem of the ctenophore Mnemiopsis leidyi invasion in the Black Sea*. 58th edn. London: International Maritime Organisation (GESAMP Reports and studies).

Ghabooli, S., Shiganova, T. A., Briski, E., Piraino, S., Fuentes, V., Thibault-Botha, D., Angel, D. L., Cristescu, M. E. and Macisaac, H. J. (2013) 'Invasion pathway of the Ctenophore Mnemiopsis leidyi in the Mediterranean Sea.', *PloS one*, 8(11), p. e81067. doi: 10.1371/journal.pone.0081067.

Ghabooli, S., Shiganova, T. A., Zhan, A., Cristescu, M. E., Eghtesadi-Araghi, P. and MacIsaac, H. J. (2011) 'Multiple introductions and invasion pathways for the invasive ctenophore Mnemiopsis leidyi in Eurasia', *Biological Invasions*, 13(3), pp. 679–690. doi: 10.1007/s10530-010-9859-8.

Gillis, N. K., Walters, L. J., Fernandes, F. C. and Hoffman, E. A. (2009) 'Higher genetic diversity in introduced than in native populations of the mussel Mytella charruana : evidence of population admixture at introduction sites', *Diversity and Distributions*, 15(5), pp. 784–795. doi: 10.1111/j.1472-

4642.2009.00591.x.

Gozlan, R. E. (2008) 'Introduction of non-native freshwater fish: is it all bad?', *Fish and Fisheries*. Blackwell Publishing Ltd, 9(1), pp. 106–115. doi: 10.1111/j.1467-2979.2007.00267.x.

Haldane, J. B. S. (1919) 'The combination of linkage values and the calculation of distances between the loci of linked factors', *Journal of Genetics*, 8, pp. 299–309. doi: 10.1016/j.biortech.2011.07.096.

Hayes, B. J., Visscher, P. M., McPartlan, H. C. and Goddard, M. E. (2003) 'Novel multilocus measure of linkage disequilibrium to estimate past effective population size', *Genome Research*, 13(4), pp. 635–643. doi: 10.1101/gr.387103.

Hewitt, C. L. and Campbell, M. L. (2007) 'Mechanisms for the prevention of marine bioinvasions for better biosecurity', *Marine Pollution Bulletin*, 55(7–9), pp. 395–401. doi: 10.1016/j.marpolbul.2007.01.005.

Hohenlohe, P. A., Day, M. D., Amish, S. J., Miller, M. R., Kamps-Hughes, N., Boyer, M. C., Muhlfeld, C. C., Allendorf, F. W., Johnson, E. A. and Luikart, G. (2013) 'Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing', *Molecular Ecology*, 22(11), pp. 3002–3013. doi: 10.1111/mec.12239.

Hufbauer, R. A., Rutschmann, A., Serrate, B., Vermeil de Conchard, H. and Facon, B. (2013) 'Role of propagule pressure in colonization success: disentangling the relative importance of demographic, genetic and habitat effects', *Journal of Evolutionary Biology*, 26(8), pp. 1691–1699. doi: 10.1111/jeb.12167.

Hulme, P. E., Bacher, S., Kenis, M., Klotz, S., Kühn, I., Minchin, D., Nentwig, W., Olenin, S., Panov, V., Pergl, J., Pyšek, P., Roques, A., Sol, D., Solarz, W. and Vilà, M. (2008) 'Grasping at the routes of biological invasions: a framework for integrating pathways into policy', *Journal of Applied Ecology*, 45(2), pp. 403–414. doi: 10.1111/j.1365-2664.2007.01442.x.

Ivanov, V. P., Kamakin, A. M., Ushivtzev, V. B., Shiganova, T., Zhukova, O., Aladin, N., Wilson, S. I., Harbison, G. R. and J. Dumont, H. (2000) 'Invasion of the Caspian Sea by the comb jellyfish Mnemiopsis leidyi (ctenophora)', *Biological Invasions*. Kluwer Academic Publishers, 2(3), pp. 255–258. doi: 10.1023/A:1010098624728.

Jaspers, C., Møller, L. F. and Kiørboe, T. (2011) 'Salinity gradient of the Baltic Sea limits the reproduction and population expansion of the newly invaded comb jelly Mnemiopsis leidyi.', *PloS one*. Edited by H. Browman, 6(8), p. e24065. doi: 10.1371/journal.pone.0024065.

Jaspers, C., Møller, L. F. and Kiørboe, T. (2015) 'Reproduction rates under variable food conditions and starvation in Mnemiopsis leidyi : significance for the invasion success of a ctenophore', *Journal of Plankton Research*, 37(5), pp. 1011–1018. doi: 10.1093/plankt/fbv017.

Javidpour, J., Sommer, U. and Shiganova, T. (2006) 'First record of Mnemiopsis leidyi A. Agassiz 1865 in the Baltic sea', *Aquatic Invasions*, 1(4), pp. 299–302. doi: 10.3391/ai.2006.1.4.17.

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M. C., White, S., Birney, E., Searle, S., Schmutz, J., Grimwood, J., Dickson, M. C., Myers, R. M., Miller, C. T., Summers, B. R., Knecht, A. K., Brady, S. D., Zhang, H., Pollen, A. a., Howes, T., Amemiya, C., Baldwin, J., Bloom, T., Jaffe, D. B., Nicol, R., Wilkinson, J., Lander, E. S., Di Palma, F., Lindblad-Toh, K. and Kingsley, D. M. (2012) 'The genomic basis of adaptive evolution in threespine sticklebacks', *Nature*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 484(7392), pp. 55–61. doi: 10.1038/nature10944.

Kamvar, Z. N. and Grunwald, N. J. (2014) 'Poppr 1 . 0 . 3 : An R package for genetic analysis of populations with mixed (clonal / sexual) reproduction', *PeerJ*, 2(e281).

Kaneps, A. G. (1979) 'Gulf stream: velocity fluctuations during the late cenozoic', *Science*, 204(4390), pp.

297–301.

Kelble, C. R., Johns, E. M., Nuttle, W. K., Lee, T. N., Smith, R. H. and Ortner, P. B. (2007) 'Salinity patterns of Florida Bay', *Estuarine, Coastal and Shelf Science*, 71(1–2), pp. 318–334. doi: 10.1016/j.ecss.2006.08.006.

Kideys, A. . and Niermann, U. (1993) 'Intrusion of Mnemiopsis mccradyi (Ctenophora: Lobata) into the Mediterranean Sea', *Senckenbergiana maritima*, 23, pp. 43–47.

Kideys, A. and Niermann, U. (1994) 'Occurrence of Mnemiopsis along the Turkish coast', *ICES Journal of Marine Science*, 51(4), pp. 423–427. doi: 10.1006/jmsc.1994.1043.

Kideys, A., Roohi, A., Bagheri, S., Finenko, G. and Kamburska, L. (2005) 'Impacts of Invasive Ctenophores on the Fisheries of the Black Sea and Caspian Sea', *Oceanography*, 18(2), pp. 76–85. doi: 10.5670/oceanog.2005.43.

Kolbe, J. J., Glor, R. E., Rodríguez Schettino, L., Lara, A. C., Larson, A. and Losos, J. B. (2004) 'Genetic variation increases during biological invasion by a Cuban lizard', *Nature*, 431(7005), pp. 177–181. doi: 10.1038/nature02807.

Konsulov, A. (1989) 'One more Newcomer. A New Species in the Black Sea-Dangerous or Harmless for Ecology', *Morski Svjat*, 2, p. 8.

Li, H. and Durbin, R. (2009) 'Fast and accurate short read alignment with Burrows-Wheeler transform', *Bioinformatics*, 25(14), pp. 1754–1760. doi: 10.1093/bioinformatics/btp324.

Lockwood, J. L., Cassey, P. and Blackburn, T. (2005) 'The role of propagule pressure in explaining species invasions', *Trends in Ecology and Evolution*, pp. 223–228. doi: 10.1016/j.tree.2005.02.004.

Lowe, S., Browne, M., Boudjelas, S. and De Poorter, M. (2000) '100 of the World's Worst Invasive Alien Species', *World*, 12(2), p. 12. doi: 10.1614/WT-04-126.1.

Luu, K., Bazin, E. and Blum, M. G. B. (2017) 'pcadapt : an R package to perform genome scans for selection based on principal component analysis', *Molecular Ecology Resources*, 17(1), pp. 67–77. doi: 10.1111/1755-0998.12592.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. and DePristo, M. A. (2010) 'The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data', *Genome Research*, 20(9), pp. 1297–1303. doi: 10.1101/gr.107524.110.

Miller, N. (1989) *Soviet relations with Latin America, 1959-1987*. Cambridge: Cambridge University Press.

Molnar, J. L., Gamboa, R. L., Revenga, C. and Spalding, M. D. (2008) 'Assessing the global threat of invasive species to marine biodiversity', *Frontiers in Ecology and the Environment*, 6(9), pp. 485–492. doi: 10.1890/070064.

Moran, E. V. and Alexander, J. M. (2014) 'Evolutionary responses to global change: Lessons from invasive species', *Ecology Letters*. Edited by F. Adler, pp. 637–649. doi: 10.1111/ele.12262.

Nei, M. and Li, W.-H. (1979) 'Mathematical model for studying genetic variation in terms of restriction endonoclueases', *Proceedings of the National Academy of Sciences of the United States of America*, 76(10), pp. 5269–5273.

Oguz, T., Fach, B. and Salihoglu, B. (2008) 'Invasion dynamics of the alien ctenophore Mnemiopsis leidyi and its impact on anchovy collapse in the Black Sea', *Journal of Plankton Research*, 30(12), pp. 1385–1397. doi: 10.1093/plankt/fbn094.

Okonechnikov, K., Conesa, A. and García-Alcalde, F. (2015) 'Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data', *Bioinformatics*, 32(2), p. btv566. doi: 10.1093/bioinformatics/btv566.

Ota, T. and Kimura, M. (1971) 'Linkage disequilibrium between two segregating nucleotide sites under the steady flux of mutations in a finite population.', *Genetics*, 68(4), pp. 571–580.

Paradis, E., Claude, J. and Strimmer, K. (2004) 'APE: Analyses of Phylogenetics and Evolution in R language', *Bioinformatics*, 20(2), pp. 289–290. doi: 10.1093/bioinformatics/btg412.

Patterson, N., Price, A. L. and Reich, D. (2006) 'Population Structure and Eigenanalysis', *PLoS Genetics*, 2(12), p. e190. doi: 10.1371/journal.pgen.0020190.

Pejchar, L. and Mooney, H. A. (2009) 'Invasive species, ecosystem services and human well-being', *Trends in Ecology & Evolution*, 24(9), pp. 497–504. doi: 10.1016/j.tree.2009.03.016.

Pereladov, M. V. (1988) 'Some observations for biota of Sudak Bay of the Black Sea', *The Third all-Russian Conference on Marine biology*, pp. 237–238.

Pimentel, D. (2011) 'Biological Invasions: Economic and Environmental Costs of Alien Plant, Animal, and Microbe Species'. CRC Press, 53(9), p. 392. doi: 10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C.

Pohlert, T. (2014) 'The Pairwise Multiple Comparison of Mean Ranks Package (PMCMR)', *R package*, p. 27. doi: http://cran.ms.unimelb.edu.au/web/packages/PMCMR/vignettes/PMCMR.pdf.

Poulos, S. E., Drakopoulos, P. G. and Collins, M. B. (1997) 'Seasonal variability in sea surface oceanographic conditions in the Aegean Sea (Eastern Mediterranean): an overview', *Journal of Marine Systems*, 13(1–4), pp. 225–244. doi: 10.1016/S0924-7963(96)00113-3.

Prachumwat, A., DeVincentis, L. and Palopoli, M. F. (2004) 'Intron Size Correlates Positively with Recombination Rate in Caenorhabditis elegans', *Genetics*, 166(3), pp. 1585–1590. doi: 10.1534/genetics.166.3.1585.

Purcell, J. E., Shiganova, T. a, Decker, M. B. and Houde, E. D. (2001) 'The ctenophore Mnemiopsis in native and exotic habitats: U.S. estuaries versus the Black Sea basin', *Hydrobiologia*, 451(1), pp. 145–176.

Purcell, J., Uye, S. and Lo, W. (2007) 'Anthropogenic causes of jellyfish blooms and their direct consequences for humans: a review', *Marine Ecology Progress Series*, 350, pp. 153–174. doi: 10.3354/meps07093.

R, core T. (2016) 'R: A language and Environment for statistical computing'.

Reeve, M. R., Syms, M. A. and Kremer, P. (1989) 'Growth dynamics of a ctenophore (Mnemiopsis) in relation to variable food supply. I. Carbon biomass, feeding, egg production, growth and assimilation efficiency', *Journal of Plankton Research*. Oxford University Press, 11(3), pp. 535–552. doi: 10.1093/plankt/11.3.535.

Reusch, T. B. H., Bolte, S., Sparwel, M., Moss, A. G. and Javidpour, J. (2010) 'Microsatellites reveal origin and genetic diversity of Eurasian invasions by one of the world's most notorious marine invader, Mnemiopsis leidyi (Ctenophora).', *Molecular ecology*, 19(13), pp. 2690–9. doi: 10.1111/j.1365-294X.2010.04701.x.

Reynolds, J., Weir, B. S. and Cockerham, C. C. (1983) 'Estimation of the coancestry coefficient: basis for a short-term genetic distance.', *Genetics*, 105(3), pp. 767–79.

Richards, C. L., Schrey, A. W. and Pigliucci, M. (2012) 'Invasion of diverse habitats by few Japanese knotweed genotypes is correlated with epigenetic differentiation', *Ecology Letters*, 15(9), pp. 1016–1025. doi: 10.1111/j.1461-0248.2012.01824.x.

Richardson, A. J., Bakun, A., Hays, G. C. and Gibbons, M. J. (2009) 'The jellyfish joyride: causes, consequences and management responses to a more gelatinous future', *Trends in Ecology & Evolution*, 24(6), pp. 312–322. doi: 10.1016/j.tree.2009.01.010.

Riisgård, H. U., Jaspers, C., Serre, S. and Lundgreen, K. (2012) 'Occurrence, inter-annual variability and zooplankton-predation impact of the invasive ctenophore Mnemiopsis leidyi and the native

jellyfish Aurelia aurita in Limfjorden (Denmark) in 2010 and 2011', *BioInvasions Records*, 1(3), pp. 145–159. doi: 10.3391/bir.2012.1.3.01.

Riquet, F., Le Cam, S., Fonteneau, E. and Viard, F. (2016) 'Moderate genetic drift is driven by extreme recruitment events in the invasive mollusk Crepidula fornicata', *Heredity*. Nature Publishing Group, 117(July 2015), pp. 1–9. doi: 10.1038/hdy.2016.24.

Riquet, F., Daguin-Thiébaut, C., Ballenghien, M., Bierne, N. and Viard, F. (2013) 'Contrasting patterns of genome-wide polymorphism in the native and invasive range of the marine mollusc Crepidula fornicata.', *Molecular ecology*, 22(4), pp. 1003–18. doi: 10.1111/mec.12161.

Rius, M., Bourne, S., Hornsby, H. G. and Chapman, M. A. (2015) 'Applications of next-generation sequencing to the study of biological invasions', *Current Zoology*, p. 488–+. doi: 10.1093/czoolo/61.3.488.

Rius, M. and Darling, J. A. (2014) 'How important is intraspecific genetic admixture to the success of colonising populations?', *Trends in ecology & evolution*, 29(4), pp. 233–42. doi: 10.1016/j.tree.2014.02.003.

Rius, M., Turon, X., Bernardi, G., Volckaert, F. and Viard, F. (2014) 'Marine invasion genetics: from spatio-temporal patterns to evolutionary outcomes', *Biological Invasions*, (in press), pp. 869–885. doi: 10.1007/s10530-014-0792-0.

Rohfritsch, A., Bierne, N., Boudry, P., Heurtebise, S., Cornette, F. and Lapègue, S. (2013) 'Population genomics shed light on the demographic and adaptive histories of European invasion in the Pacific oyster, Crassostrea gigas.', *Evolutionary applications*, 6(7), pp. 1064–78. doi: 10.1111/eva.12086.

Roman, J. and Darling, J. a. (2007) 'Paradox lost: genetic diversity and the success of aquatic invasions', *Trends in Ecology and Evolution*, 22(9), pp. 454–464. doi: 10.1016/j.tree.2007.07.002.

Roohi, A., Kideys, A. E., Sajjadi, A., Hashemian, A., Pourgholam, R., Fazli, H., Khanari, A. G. and Eker-Develi, E. (2010) 'Changes in biodiversity of phytoplankton, zooplankton, fishes and macrobenthos in the Southern Caspian Sea after the invasion of the ctenophore Mnemiopsis Leidyi', *Biological Invasions*, 12(7), pp. 2343–2361. doi: 10.1007/s10530-009-9648-4.

Ryan, J. F., Pang, K., Schnitzler, C. E., Nguyen, A.-D., Moreland, R. T., Simmons, D. K., Koch, B. J., Francis, W. R., Havlak, P., Smith, S. A., Putnam, N. H., Haddock, S. H. D., Dunn, C. W., Wolfsberg, T. G., Mullikin, J. C., Martindale, M. Q. and Baxevanis, A. D. (2013) 'The genome of the ctenophore Mnemiopsis leidyi and its implications for cell type evolution.', *Science (New York, N.Y.)*, 342(6164), p. 1242592. doi: 10.1126/science.1242592.

Sasson, D. A. and Ryan, J. F. (2016) 'The sex lives of ctenophores: the influence of light, body size, and self-fertilization on the reproductive output of the sea walnut, *Mnemiopsis leidyi*', *PeerJ*. PeerJ, Inc, 4, p. e1846. doi: 10.7717/peerj.1846.

Sax, D. F. and Brown, J. H. (2000) 'The paradox of invasion', *Global Ecology and Biogeography*, 9(5), pp. 363–371. doi: 10.1046/j.1365-2699.2000.00217.x.

Schlaepfer, M. A., Sax, D. F. and Olden, J. D. (2011) 'The potential conservation value of non-native species.', *Conservation biology : the journal of the Society for Conservation Biology*. Blackwell Publishing Inc, 25(3), pp. 428–37. doi: 10.1111/j.1523-1739.2010.01646.x.

Seebens, H., Gastner, M. T. and Blasius, B. (2013) 'The risk of marine bioinvasion caused by global shipping', *Ecology Letters*. Edited by Franck Courchamp, 16(6), pp. 782–790. doi: 10.1111/ele.12111.

Simberloff, D. (2009) 'The Role of Propagule Pressure in Biological Invasions', *Annual Review of Ecology Evolution and Systematics*, 40(1), pp. 81–102. doi: 10.1146/annurev.ecolsys.110308.120304.

Simberloff, D., Martin, J.-L., Genovesi, P., Maris, V., Wardle, D. A., Aronson, J., Courchamp, F., Galil, B., García-Berthou, E., Pascal, M., Pyšek, P., Sousa, R., Tabacchi, E. and Vilà, M. (2013) 'Impacts of biological

invasions: what's what and the way forward', *Trends in Ecology & Evolution*, 28(1), pp. 58–66. doi: 10.1016/j.tree.2012.07.013.

Sorte, C. J. B., Ibáñez, I., Blumenthal, D. M., Molinari, N. A., Miller, L. P., Grosholz, E. D., Diez, J. M., D'Antonio, C. M., Olden, J. D., Jones, S. J. and Dukes, J. S. (2013) 'Poised to prosper? A cross-system comparison of climate change effects on native and non-native species performance', *Ecology Letters*. Edited by K. Suding, 16(2), pp. 261–270. doi: 10.1111/ele.12017.

Stanev, E. V. (1990) 'On the mechanisms of the Black Sea circulation', *Earth-Science Reviews*, 28(4), pp. 285–319. doi: 10.1016/0012-8252(90)90052-W.

Stopar, K., Ramšak, A., Trontelj, P. and Malej, A. (2010) 'Lack of genetic structure in the jellyfish Pelagia noctiluca (Cnidaria: Scyphozoa: Semaeostomeae) across European seas', *Molecular Phylogenetics and Evolution*, 57(1), pp. 417–428. doi: 10.1016/j.ympev.2010.07.004.

Sved, J. A. and Feldman, M. W. (1973) 'Correlation and probability methods for one and two loci.', *Theoretical population biology*, 4(1), pp. 129–32. doi: 10.1016/0040-5809(73)90008-7.

Szűcs, M., Melbourne, B. a, Tuff, T. and Hufbauer, R. a (2014) 'The roles of demography and genetics in the early stages of colonization.', *Proceedings. Biological sciences / The Royal Society*, 281(1792), p. 20141073-. doi: 10.1098/rspb.2014.1073.

Tajima, F. (1989) 'Statistical method for testing the neutral mutation hypothesis by DNA polymorphism.', *Genetics*, 123(3), pp. 585–95. doi: PMC1203831.

Tatem, A. J., Hay, S. I. and Rogers, D. J. (2006) 'Global traffic and disease vector dispersal.', *Proceedings of the National Academy of Sciences of the United States of America*, 103(16), pp. 6242–6247. doi: 10.1073/pnas.0508391103.

Taylor, C. M. and Hastings, A. (2005) 'Allee effects in biological invasions', *Ecology Letters*. Blackwell Science Ltd, 8(8), pp. 895–908. doi: 10.1111/j.1461-0248.2005.00787.x.

Tenesa, A., Navarro, P., Hayes, B. J., Duffy, D. L., Clarke, G. M., Goddard, M. E. and Visscher, P. M. (2007) 'Recent human effective population size estimated from linkage disequilibrium', *Genome Research*, 17(4), pp. 520–526. doi: 10.1101/gr.6023607.

Tepolt, C. K. (2015) 'Adaptation in marine invasion: a genetic perspective', *Biological Invasions*, 17(3), pp. 887–903. doi: 10.1007/s10530-014-0825-8.

Tepolt, C. K. and Palumbi, S. R. (2015) 'Transcriptome sequencing reveals both neutral and adaptive genome dynamics in a marine invader.', *Molecular ecology*, 24(16), pp. 4145–58. doi: 10.1111/mec.13294.

Vera, M., Díez-Del-Molino, D. and García-Marín, J. L. (2016) 'Genomic survey provides insights into the evolutionary changes that occurred during European expansion of the invasive mosquitofish (Gambusia holbrooki)', *Molecular Ecology*, 25(5), pp. 1089–1105. doi: 10.1111/mec.13545.

Verhoeven, K. J. F., Macel, M., Wolfe, L. M. and Biere, A. (2011) 'Population admixture, biological invasions and the balance between local adaptation and inbreeding depression.', *Proceedings. Biological sciences / The Royal Society*, 278(1702), pp. 2–8. doi: 10.1098/rspb.2010.1272.

Viard, F., David, P. and Darling, J. A. (2016) 'Marine invasions enter the genomic era: three lessons from the past, and the way forward', *Current Zoology*. The Oxford University Press, p. zow053. doi: 10.1093/cz/zow053.

Vinogradov, M. Y., Shushkina, E. A., Musayeva, E. I. and Sorokin, P. Y. (1989) 'A newly acclimated species in the Black Sea: the ctenophore Mnemiopsis leidyi (Ctenophora: Lobata)', *Oceanology*, 29(2).

van Walraven, L., Langenberg, V. T. and Van der Veer, H. W. (2013) 'Seasonal occurrence of the invasive ctenophore Mnemiopsis leidyi in the western Dutch Wadden Sea', *Journal of Sea Research*, 82, pp. 86–92. doi: 10.1016/j.seares.2013.02.003.

Watterson, G. A. (1975) 'On the number of segregating sites in genetical models without recombination', *Theoretical Population Biology*, 7(2), pp. 256–276. doi: 10.1016/0040-5809(75)90020-9.

Weir, B. S. and Cockerham, C. C. (1984) 'Estimating F-Statistics for the Analysis of Population Structure', *Evolution*, 38(6), p. 1358. doi: 10.2307/2408641.

Whitney, K. D. and Gering, E. (2015) 'Five decades of invasion genetics.', *The New phytologist*, 205(2), pp. 472–5. doi: 10.1111/nph.13197.

Wigginton, J. E., Cutler, D. J. and Abecasis, G. R. (2005) 'A Note on Exact Tests of Hardy-Weinberg Equilibrium', *The American Journal of Human Genetics*, 76(5), pp. 887–893. doi: 10.1086/429864.

Williamson, M. and Fitter, A. (1996) 'The Varying Success of Invaders', *Ecology*, 77(6), pp. 1661–1666. doi: 10.2307/2265769.

Willing, E.-M., Dreyer, C. and van Oosterhout, C. (2012) 'Estimates of Genetic Differentiation Measured by FST Do Not Necessarily Require Large Sample Sizes When Using Many SNP Markers', *PLoS ONE*. Edited by J. Chave. Public Library of Science, 7(8), p. e42649. doi: 10.1371/journal.pone.0042649.

De Wit, P. and Palumbi, S. R. (2013) 'Transcriptome-wide polymorphisms of red abalone (Haliotis rufescens) reveal patterns of gene flow and local adaptation', *Molecular Ecology*, 22(11), pp. 2884–2897. doi: 10.1111/mec.12081.

Wonham, M. J., Walton, W. C., Ruiz, G. M., Frese, A. M. and Galil, B. S. (2001) 'Going to the source: Role of the invasion pathway in determining potential invaders', *Marine Ecology Progress Series*, 215, pp. 1–12. doi: 10.3354/meps215001.

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., Goddard, M. E. and Visscher, P. M. (2010) 'Common SNPs explain a large proportion of the heritability for human height.', *Nature genetics*, 42(7), pp. 565–9. doi: 10.1038/ng.608.

Yu, N., Chen, F.-C., Ota, S., Jorde, L. B., Pamilo, P., Patthy, L., Ramsay, M., Jenkins, T., Shyue, S.-K. and Li, W.-H. (2002) 'Larger genetic differences within africans than between Africans and Eurasians.', *Genetics*, 161(1), pp. 269–74.

Supplements

Supplement A

Table A: Sampling locations and coordinates including sample size and first records of *M. leidyi* in each region.

Location	*Region*	*Latitude (deg)*	*Longitude (deg)*	*First Record*	*n*
Woods Hole, US	*New England*	*41.5225234*	*-070.6744909*	*-*	*16*
Miami, US	*Floridian Peninsula*	*25.6146556*	*-080.3053167*	*-*	*15*
Sylt, DE	*North Sea*	*55.0169422*	*008.4402412*	*2006* [a]	*16*
Varna, BG	*Western Black Sea*	*43.1839869*	*027.9000270*	*1986* [b]	*16*
Villefranche-sur-Mer, FR	*Western Mediterranean*	*43.6884186*	*007.3148346*	*2009* [c]	*11*

[a] Faasse and Bayha (2006)

[b] Konsulov (1989)

[c] Fuentes *et al.* (2009) & Boero *et al.* (2009)

Supplement B

M. leidyi Lyophilization Protocol

1. Briefly rinse specimen in freshwater to minimise salt carryover. Check for solid contaminants and remove them if possible.

2. Transfer specimen into a suitable airtight container (snap cap vial, Falcon tube, Ziploc bag etc.). Ensure the specimen occupies **less than one third of the total volume**. Samples tend to foam during lyophilization and may cross-contaminate if they "bubble-over".

3. Pre-freeze samples at **-80 °C** for **>2h**. Storage at -20 °C is not recommended as this is close to the eutectic point of seawater and may lead to incomplete lyophilization.

4. Prepare the freeze dryer according to manufacturer's instructions. Open the sample containers and add the pre-frozen samples to the condensation chamber once this has reached a stable temperature of **-50 °C** or below. Ensure that sample containers are stood upright and cross-contamination cannot occur.

5. Freeze-dry the samples for ~**48h**. After lyophilization samples should have turned into a **visibly dry** and **odourless** white substance. If there are signs of incomplete drying, continue the process for a further 6-12h and check again.

6. Close the sample containers and store the samples at **RT** in a **dry environment**. Avoid opening and closing the containers after lyophilization as this will introduce moisture. Samples can be stored alongside silica gel to ensure dry conditions.

Supplement C

M. leidyi High-Molecular-Weight DNA Extraction Protocol (CTAB/Chloroform Method)

Required Reagents:

Lysis-Extraction-Buffer (see below)
Ethanol 70% (4°C)
ß-mercaptoethanol
Isopropanol 100% (4°C)
Chloroform-Isoamyl-Alcohol 24:1 vol:vol
Proteinase K
TElow-Buffer (10mM Tris-HCl & 0.1mM EDTA)

OPTIONAL:
RNAse A
Sodiumacetate (3M)

Per **1 mL Lysis-Extraction-Buffer** (approx. 1 sample) add:

CTAB		20 mg
NaCl	0.0014 Mol =	82 mg
EDTA (0.5M)		40 µl
Tris-HCl (1M pH 8)		100 µl
H_2O		fill up to 1 ml

The solution does not need to be autoclaved and can be stored for several weeks at RT. **Renew EDTA und Tris-HCl stocks every 2 months.**

1 Add **0.5-1 cm³** of lyophilized tissue to a **2 mL** tube. **Ensure flakes of tissue do not contaminate other tubes (close these!).**

2 Clean dissection instruments with **70% Ethanol** between each sample. Dry with clean wipes.

3 Submerge tissue in **1 mL LE-Buffer** (add up to 1.5 ml if necessary). Add **Proteinase K (1 mg/ml)** and vortex briefly. Break up large pieces of tissue with a clean pipette tip.

4 ***OPTIONAL RNA DIGEST*** (If not required go to step 5): Add **RNAse A (100 µg/ml)** to the suspension and vortex briefly. Incubate samples in the thermoshaker for **30 min** at **37°C**.

5 Add **2 µl** of **ß-mercaptoethanol** and lyse samples for **1-2h** in the thermoshaker at **56°C** until a homogenous solution remains. Intermittently mix samples by inverting the tubes. Small pieces of tissue that remain are not of concern and will be centrifuged in the next step.

6 Add **800 µl** of **Chloroform:Isoamylalcohol** and mix samples for **2-3 min** by gently inverting the tubes until a homogenous, milky emulsion forms. IMPORTANT: Do not shake or vortex!

7 Centrifuge for **5 min** at **7500 x g**. Remove the **aqueous phase (Top Layer)** using a pipette and transfer it to a new **2 ml Tube**. **Avoid contact with the protein interphase (Central solid layer) and the organic phase (Lower layer)!** Discard inter- and organic phase (Toxic waste!).

8 Repeat **steps 6** and **7** and transfer the **aqueous phase** into a new **1.5 ml Tube**. The yield should be approx. **800–900 µl**.

9 Add **2/3 of the Volume** of the aqueous phase (approx. 500-600 µl) of **cold (4°C) Isopropanol**, mix by inversion. Work on ice from now on.

10 Precipitate DNA for **10 min** at **4°C**.

11 Centrifuge for **10 min** at **max. RPM**, note the orientation of the pellet. Pipet off or pour out supernatant (Do not lose the pellet!). Dry tube rim using clean wipe.

12 Wash pellet with **1 ml cold (4°C) 70% EtOH**, gently invert to detach the pellet from the tube, and wash for **2 min**.

13 Centrifuge for **3 min** at **max. RPM** to reattach pellet. Remove ethanol with a pipette and dry pellet for **10 min** at **RT**. Ensure ethanol has completely evaporated before proceeding.

14 **OPTIONAL:** Repeat ethanol wash, especially if salt residues remain.

15 Depending on the required concentration, dissolve pellet in **20-200 µl** of **TElow-Buffer** at **RT** whilst swirling gently. Store at **max. –20°C**.

Supplement D

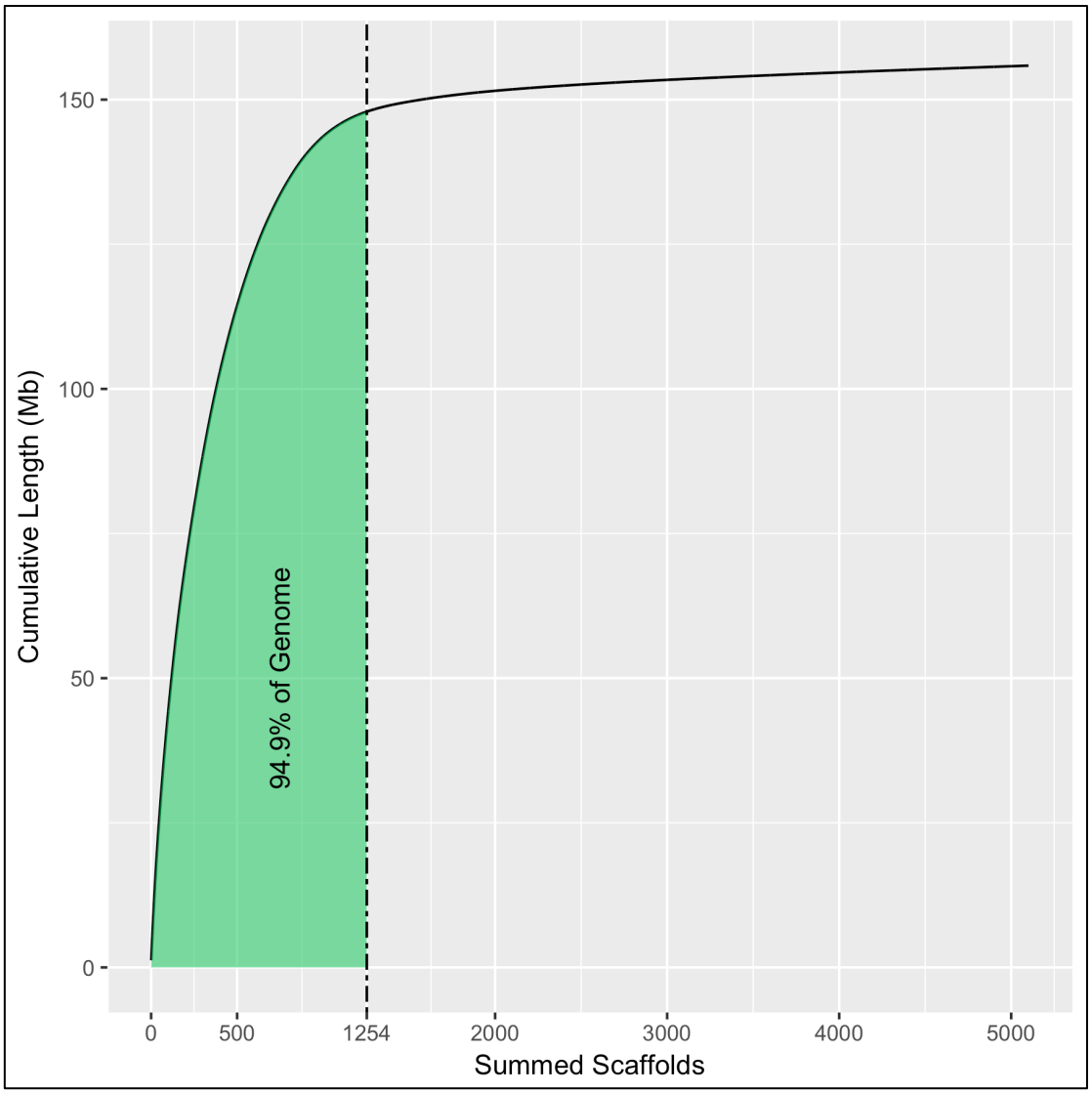

Figure D: Cumulative length of scaffolds sorted by size. Dashed line marks the smallest scaffold with length >10 Kb. Green shaded area shows scaffolds retained for analysis containing 94.9% of the total sequence.

Supplement E

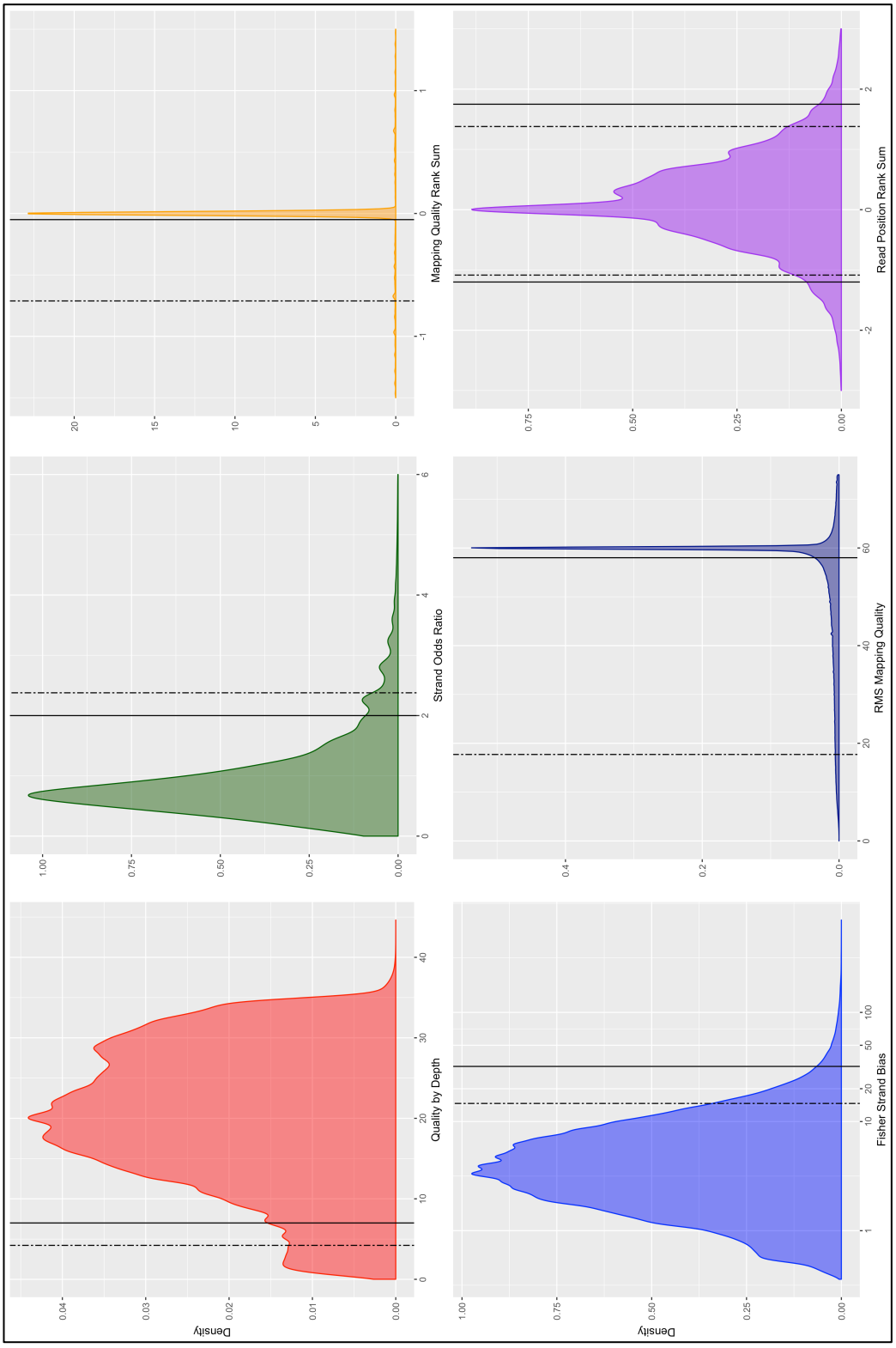

Figure E: Distributions of SNP quality metrics after initial variant calling. Solid lines mark the hard-filter thresholds applied. Dashed lines mark the 5th and/or 95th quantile respectively for reference purposes.

Supplement F

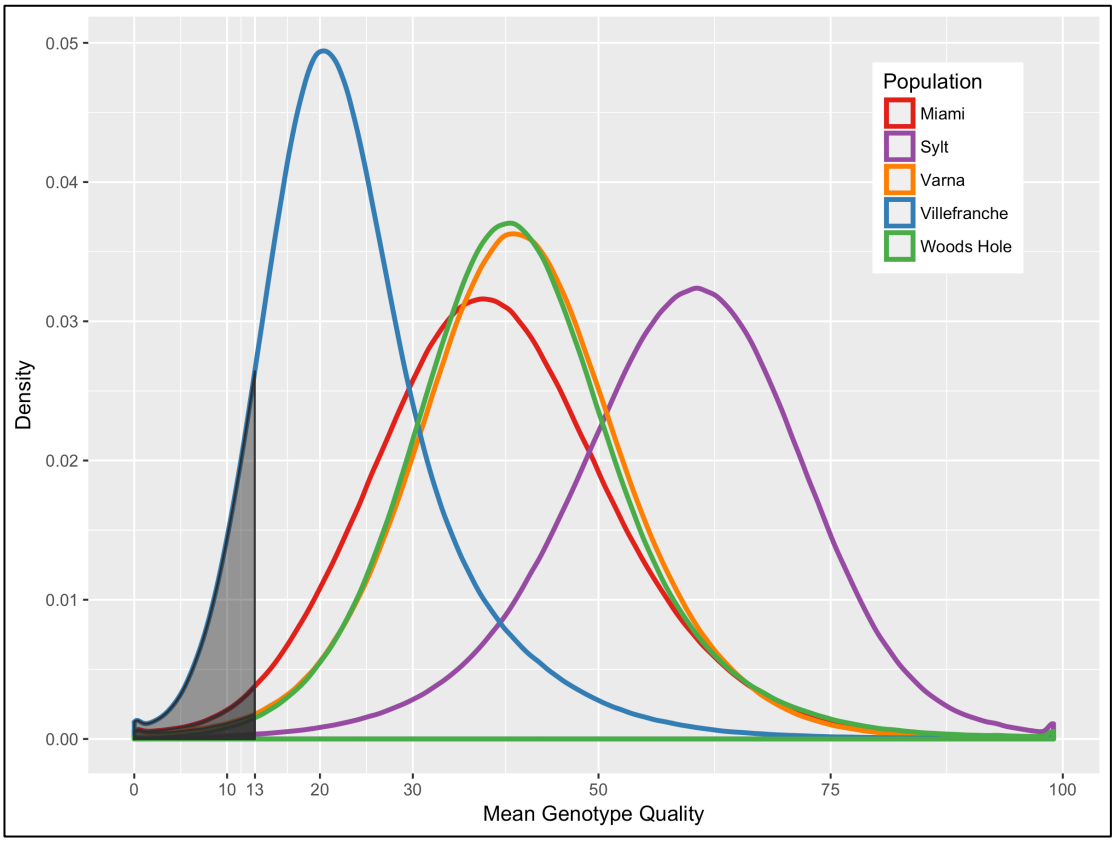

Figure F: Distributions of genotype qualities by population. Grey shaded area marks low-quality, discarded genotypes.

Supplement G

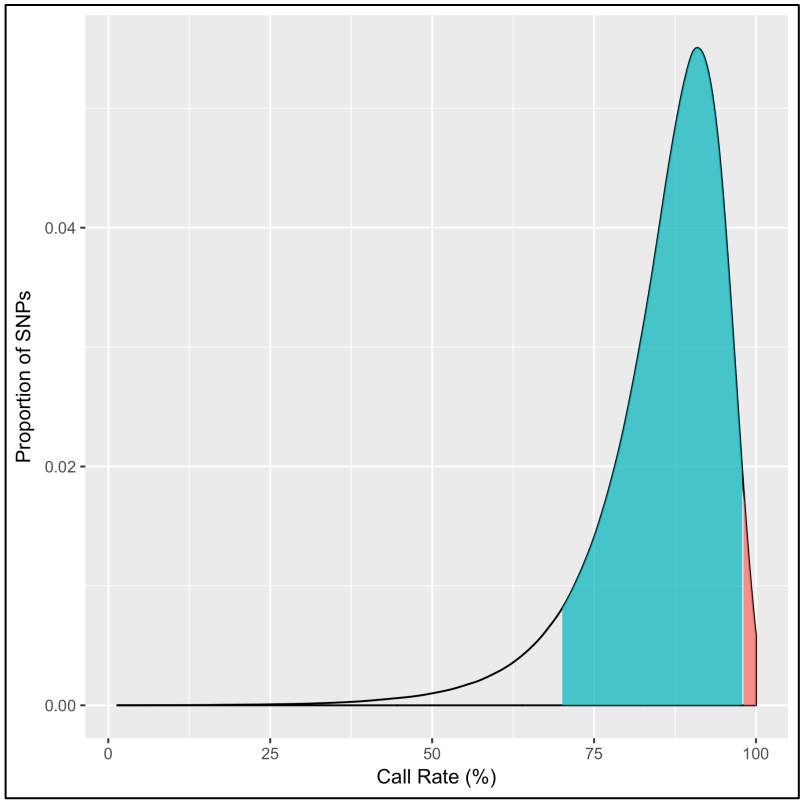

Figure G-1: Distribution of global SNP call rates. A 100% call rate means every individual was genotyped at that locus. The blue shaded area marks SNPs with >70% call rate, used in the lenient SNP set. The red shaded area marks SNPs where at most 1 genotype is missing, used in the strict SNP set.

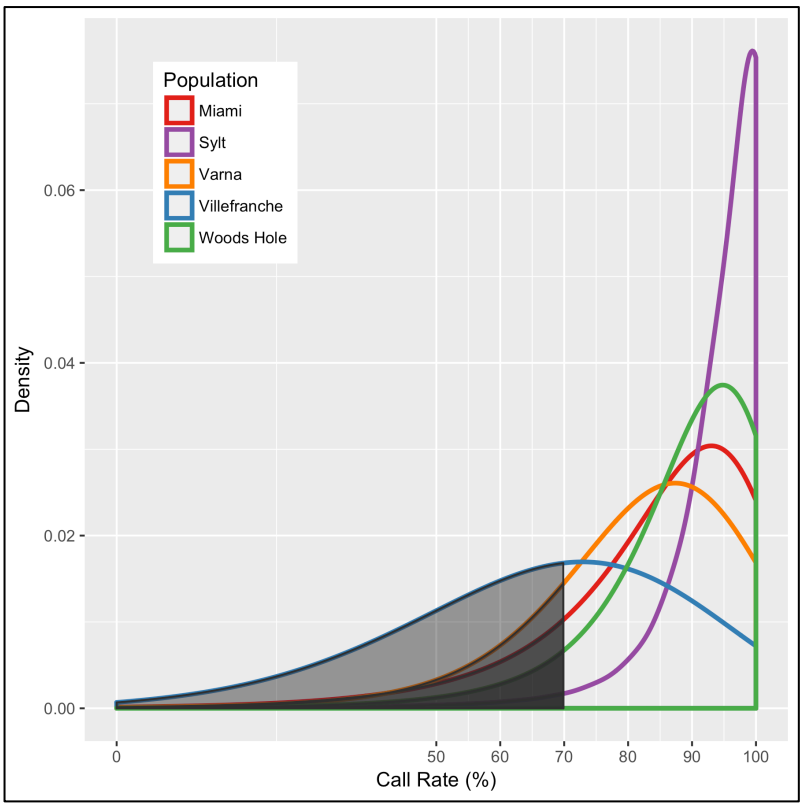

Figure G-2: Distribution of SNP call rates by population. A 100% call rate means every individual in a population has been genotyped at that locus. Grey shaded area marks discarded SNPs with call rates below 70%.

Supplement H

Table H: Sequencing output and mean base qualities of the six NextSeq 500 flowcells.

Flowcell	*Output (Gb)*	*Mean Base Quality*
1	*29.3*	*27.6*
2	*30.2*	*26.1*
3	*29.7*	*27.9*
4	*30.4*	*26.2*
5	*126.8*	*29.5*
6	*133.9*	*28.6*
Total	***380.299***	***28.4***

Supplement I

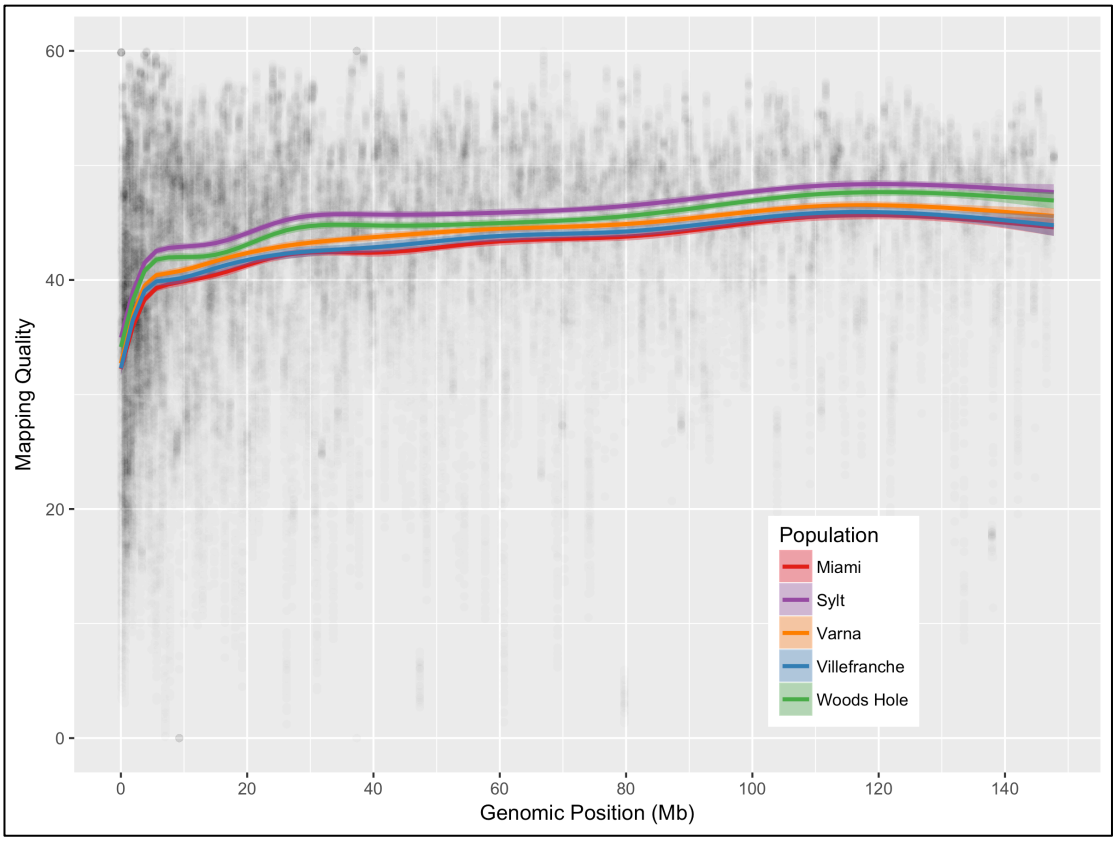

Figure I-1: Phred-scaled mean mapping quality across concatenated scaffolds. Black transparent dots represent point-means every 10 Kb. Coloured lines are running means for each population. Shaded areas around the lines represent the 95% confidence interval.

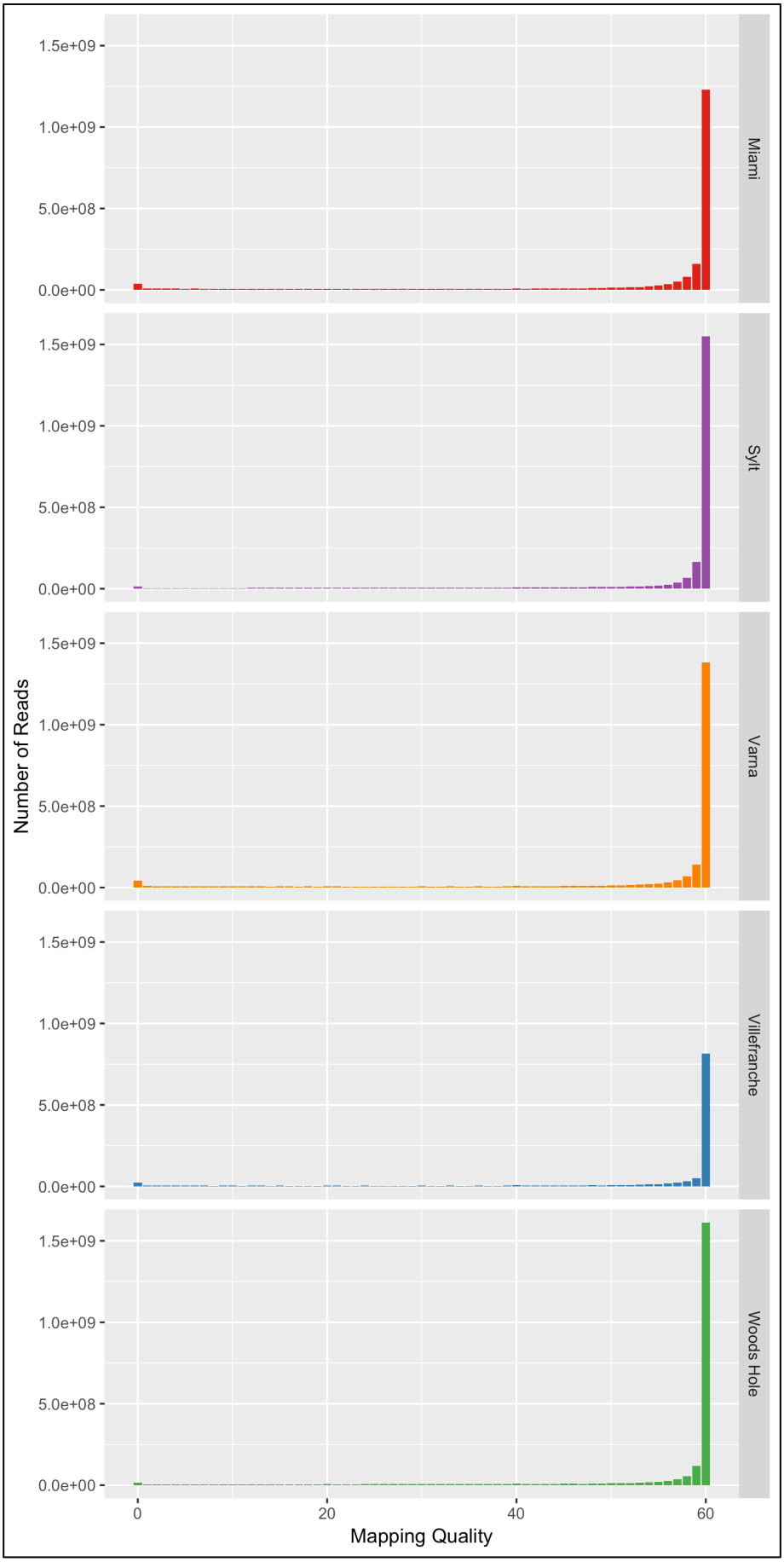

Figure I-2: Number of reads of a given mapping quality split by population. A mapping quality of zero implies either non-mapping or ambiguously mapping reads.

Supplement J

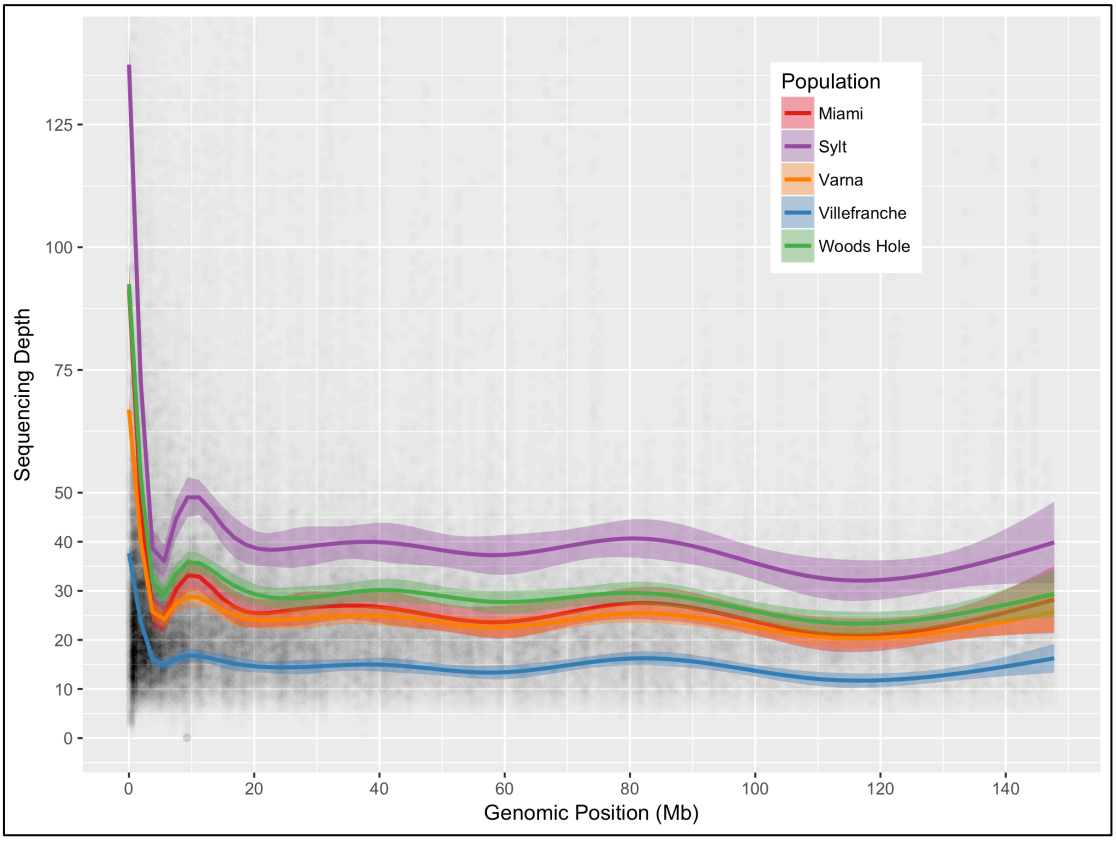

Figure J-1: Mean sequencing depth across concatenated scaffolds. Transparent, black dots display point means every 10 Kb. Coloured lines show the running means per population. Shaded areas around the lines mark the 95% confidence interval.

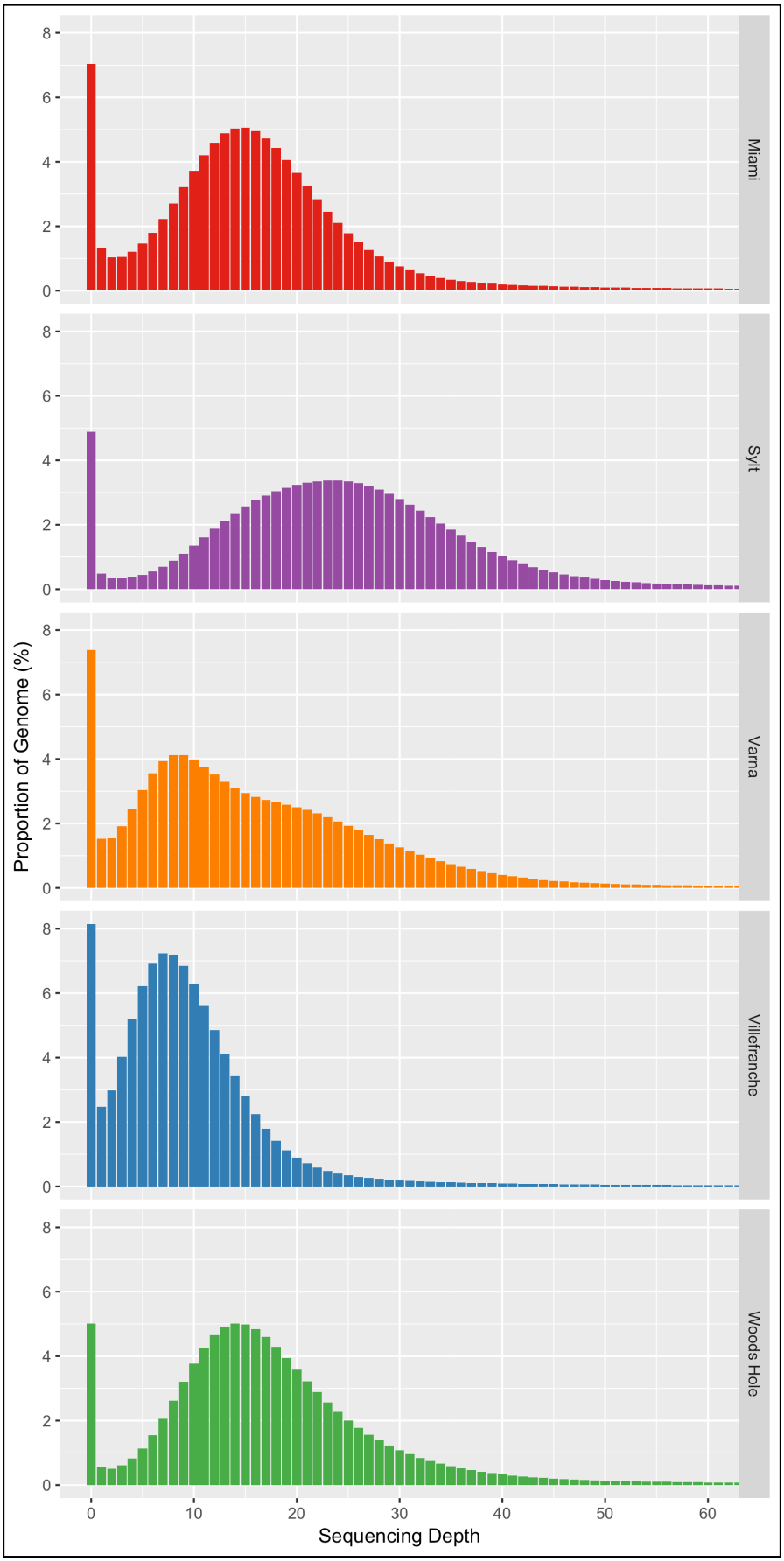

Figure J-2: Proportion of genome with a given mean sequencing depth for each population. The zero-bin denotes the proportion of the genome that was not sequenced and/or no reads were mapped. The horizontal axis has been truncated for easier visualisation.

Supplement K

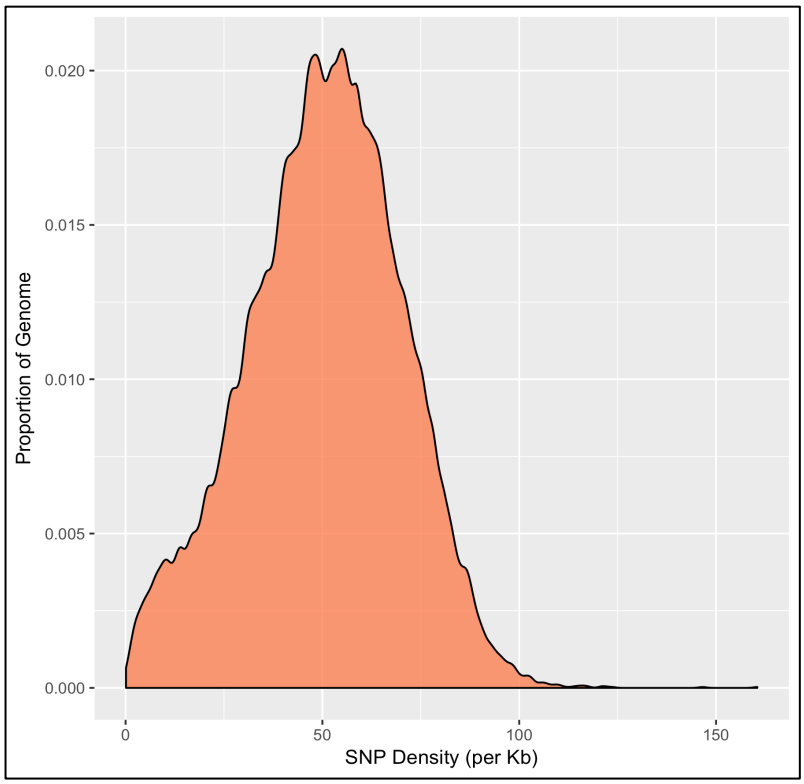

Figure K: Genome-wide distribution of SNP density per Kb.

Supplement L

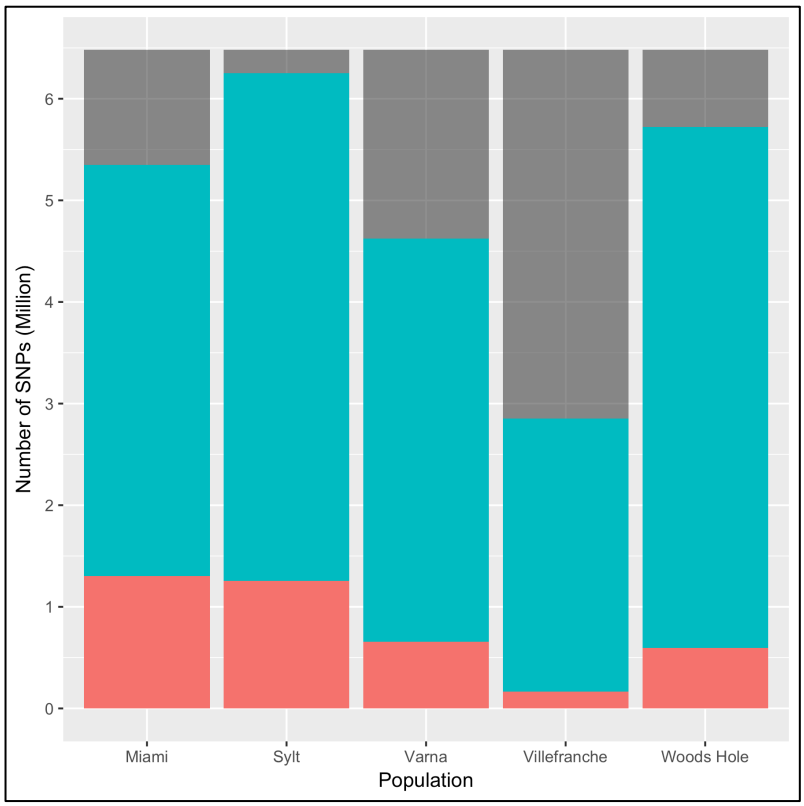

Figure L: Population-specific SNP counts for the lenient (blue) and strict (red) sets after subsampling to the largest common sample size. Grey bars show the global count.

Supplement M

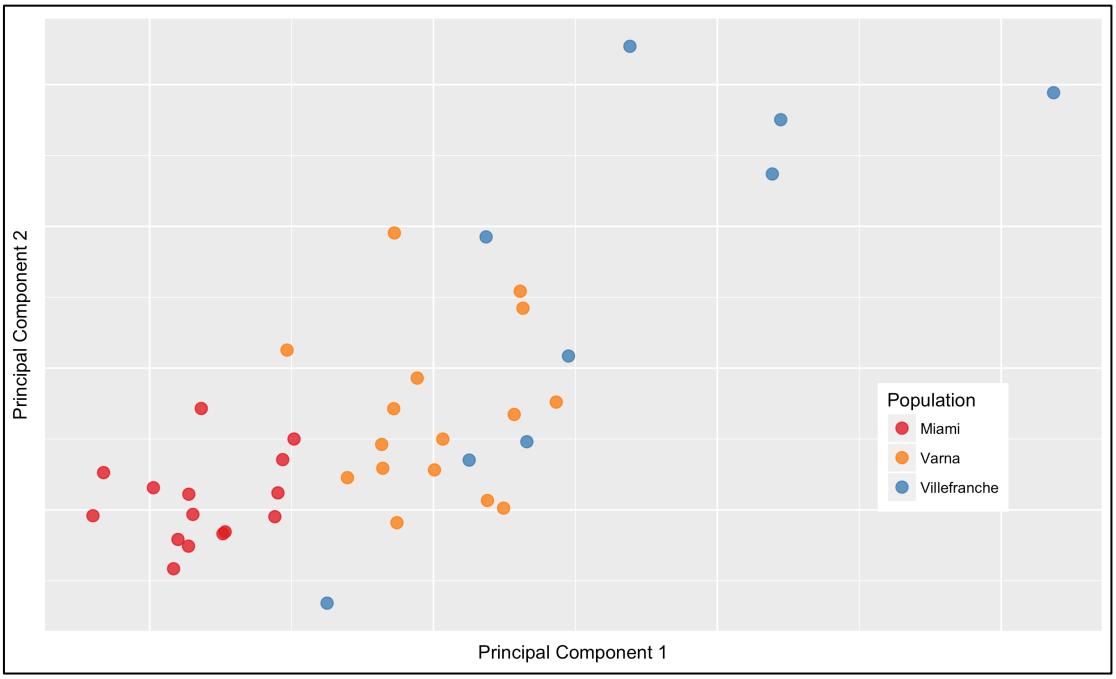

Figure M: Enlarged and rotated section of Figure 8. Zoom into southern individuals.

Supplement N

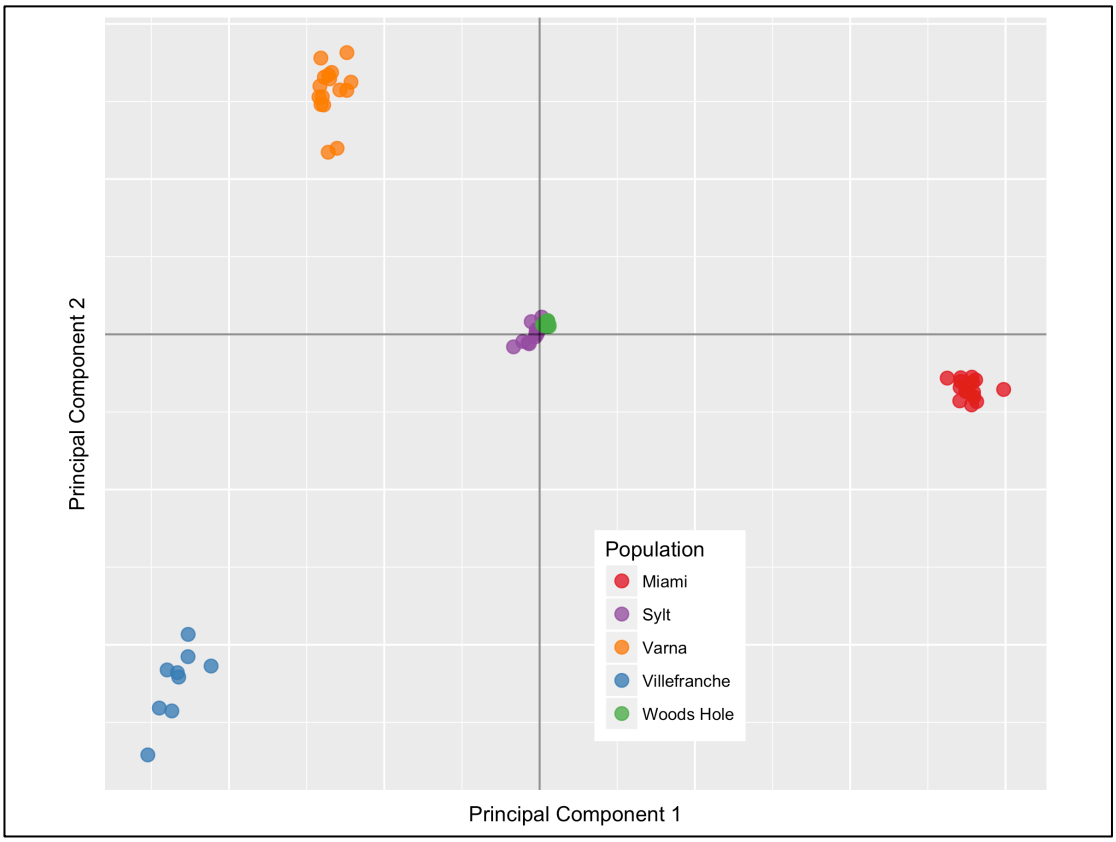

Figure N: Northern populations projected onto the principal components that were independently derived from a PCA of the southern populations.

Supplement O

Table O: Summary of one-sample t-tests for each weighted and unweighted pairwise F_{ST}.

Weighted Pairwise F_{ST}					
Pairwise Comparison	***t***	***D.o.F.***	***p***	***C.I. (95%)***	
Miami : Varna	*510*	*135760*	***<< 0.001***	*0.100*	*0.101*
Miami : Villefranche	*607*	*135720*	***<< 0.001***	*0.146*	*0.147*
Sylt : Miami	*1068*	*135840*	***<< 0.001***	*0.409*	*0.411*
Sylt : Varna	*1071*	*135820*	***<< 0.001***	*0.401*	*0.402*
Sylt : Villefranche	*1013*	*135820*	***<< 0.001***	*0.458*	*0.460*
Sylt : Woods Hole	*170*	*135880*	***<< 0.001***	*0.020*	*0.021*
Varna : Villefranche	*469*	*135660*	***<< 0.001***	*0.069*	*0.069*
Woods Hole : Miami	*1115*	*135840*	***<< 0.001***	*0.433*	*0.435*
Woods Hole : Varna	*1120*	*135800*	***<< 0.001***	*0.426*	*0.428*
Woods Hole : Villefranche	*1063*	*135800*	***<< 0.001***	*0.492*	*0.494*

Unweighted Pairwise F_{ST}					
Pairwise Comparison	***t***	***D.o.F.***	***p***	***C.I. (95%)***	
Miami : Varna	*637*	*135760*	***<< 0.001***	*0.058*	*0.058*
Miami : Villefranche	*658*	*135720*	***<< 0.001***	*0.084*	*0.085*
Sylt : Miami	*1118*	*135840*	***<< 0.001***	*0.210*	*0.211*
Sylt : Varna	*1125*	*135820*	***<< 0.001***	*0.202*	*0.203*
Sylt : Villefranche	*1057*	*135820*	***<< 0.001***	*0.298*	*0.299*
Sylt : Woods Hole	*181*	*135880*	***<< 0.001***	*0.016*	*0.016*
Varna : Villefranche	*482*	*135660*	***<< 0.001***	*0.038*	*0.038*
Woods Hole : Miami	*1145*	*135840*	***<< 0.001***	*0.220*	*0.221*
Woods Hole : Varna	*1143*	*135800*	***<< 0.001***	*0.211*	*0.212*
Woods Hole : Villefranche	*1104*	*135800*	***<< 0.001***	*0.316*	*0.317*

Supplement P

Table P-1: Summary statistics of Games-Howell post-hoc tests for Individual Multilocus Heterozygosity (IMLH) following a significant (p << 0.001) Welch-corrected ANOVA. Significant *p*-values have been highlighted in bold.

Comparison	*Mean Diff.*	*S.E.*	*t*	*D.F.*	*p*	*C.I. (95%)*	
Sylt : Woods Hole	*0.012*	*0.005*	*1.90*	*11.4*	*0.372*	*0.033*	*-0.009*
Sylt : Miami	*-0.030*	*0.004*	*4.90*	*9.8*	***0.005***	*-0.010*	*-0.051*
Sylt : Varna	*-0.035*	*0.005*	*4.50*	*15.6*	***0.003***	*-0.011*	*-0.059*
Sylt : Villefranche	*0.000*	*0.005*	*0.05*	*13.2*	*1.000*	*0.021*	*-0.022*
Woods Hole : Miami	*-0.043*	*0.002*	*12.48*	*14.5*	***<0.001***	*-0.032*	*-0.053*
Woods Hole : Varna	*-0.047*	*0.004*	*8.24*	*12.5*	***<0.001***	*-0.029*	*-0.065*
Woods Hole : Villefranche	*-0.013*	*0.003*	*2.80*	*15.1*	*0.084*	*0.001*	*-0.027*
Miami : Varna	*-0.004*	*0.004*	*0.80*	*10.5*	*0.924*	*0.013*	*-0.022*
Miami : Villefranche	*0.030*	*0.003*	*7.33*	*12.5*	***<0.001***	*0.043*	*0.017*
Varna : Villefranche	*0.034*	*0.004*	*5.60*	*14.5*	***<0.001***	*0.054*	*0.015*

Table P-2: Summary statistics of Nemenyi post-hoc tests with Chi-square approximation for population-level heterozygosity (H_S) following a significant (p << 0.001) Kruskal-Wallis omnibus test. Significant *p*-values have been highlighted in bold.

	Miami	*Sylt*	*Varna*	*Villefranche*
Sylt	***<<0.001***			
Varna	*0.260*	***<<0.001***		
Villefranche	***<<0.001***	***<<0.001***	***<<0.001***	
Woods Hole	***<<0.001***	***<<0.001***	***<<0.001***	*0.300*

Supplement Q

Table Q: Summary statistics of Nemenyi post-hoc tests with Chi-square approximation for genome-wide distributions of Watterson's theta (θ_W) following a significant (p << 0.001) Kruskal-Wallis omnibus test. Significant *p*-values have been highlighted in bold.

	Miami	***Sylt***	***Varna***	***Villefranche***
Sylt	***<<0.001***			
Varna	*0.440*	***<<0.001***		
Villefranche	***<<0.001***	***<<0.001***	***<<0.001***	
Woods Hole	***<<0.001***	***<<0.001***	***<<0.001***	***<<0.001***

Supplement R

Table R: Summary statistics of Nemenyi post-hoc tests with Chi-square approximation for genome-wide distributions of Nei's Nucleotide diversity (π) following a significant (p << 0.001) Kruskal-Wallis omnibus test. Significant *p*-values have been highlighted in bold.

	Miami	***Sylt***	***Varna***	***Villefranche***
Sylt	***<<0.001***			
Varna	***<<0.001***	***<<0.001***		
Villefranche	***<<0.001***	***<<0.001***	***<<0.001***	
Woods Hole	***<<0.001***	***<<0.001***	***<<0.001***	***<<0.001***

Supplement S

Table S-1: Summary statistics of Games-Howell post-hoc tests for individual inbreeding coefficients (*F*) following a significant (p < 0.001) Welch-corrected ANOVA. Significant *p*-values have been highlighted in bold.

Comparison	*Mean Diff.*	*S.E.*	*t*	*D.F.*	*p*	*C.I. (95%)*	
Sylt : Woods Hole	*-0.005*	*0.011*	*0.33*	*20.2*	*0.997*	*0.043*	*-0.054*
Sylt : Miami	*0.039*	*0.011*	*2.49*	*19.0*	*0.135*	*0.087*	*-0.008*
Sylt : Varna	*0.032*	*0.014*	*1.56*	*29.8*	*0.531*	*0.09*	*-0.027*
Sylt : Villefranche	*0.040*	*0.014*	*2.10*	*22.8*	*0.254*	*0.097*	*-0.016*
Woods Hole : Miami	*0.045*	*0.006*	*5.35*	*28.7*	***<0.001***	*0.069*	*0.020*
Woods Hole : Varna	*0.037*	*0.011*	*2.45*	*21.1*	*0.140*	*0.082*	*-0.008*
Woods Hole : Villefranche	*0.045*	*0.010*	*3.35*	*12.5*	***0.036***	*0.088*	*0.002*
Miami : Varna	*-0.008*	*0.010*	*0.53*	*19.7*	*0.983*	*0.036*	*-0.052*
Miami : Villefranche	*0.001*	*0.009*	*0.05*	*11.4*	*1.000*	*0.043*	*-0.042*
Varna : Villefranche	*0.009*	*0.013*	*0.47*	*22.2*	*0.989*	*0.063*	*-0.045*

Table S-1: Summary of one-sample t-tests for individual inbreeding coefficient (*F*) for each population.

Population	*t*	*D.o.F.*	*p*	*C.I. (95%)*	
Miami	*2.52*	*14*	***0.025***	*0.002*	*0.026*
Sylt	*-1.72*	*15*	*0.106*	*-0.057*	*0.006*
Varna	*0.44*	*15*	*0.665*	*-0.023*	*0.035*
Villefranche	*1.21*	*8*	*0.260*	*-0.013*	*0.042*
Woods Hole	*-4.90*	*15*	***<0.001***	*-0.044*	*-0.017*

Supplement T

Table T-1: Summary statistics of Games-Howell post-hoc tests for Tajima's *D* following a significant (p << 0.001) Welch-corrected ANOVA. Significant *p*-values have been highlighted in bold.

Comparison	*Mean Diff.*	*S.E.*	*t*	*D.F.*	*p*	*C.I. (95%)*	
Sylt : Woods Hole	*0.244*	*0.011*	*15.1*	*28277*	***<0.001***	*0.288*	*0.200*
Sylt : Miami	*-0.232*	*0.009*	*19.2*	*18598*	***<0.001***	*-0.199*	*-0.264*
Sylt : Varna	*-0.287*	*0.008*	*23.9*	*18241*	***<0.001***	*-0.254*	*-0.319*
Sylt : Villefranche	*0.212*	*0.009*	*17.2*	*20103*	***<0.001***	*0.246*	*0.179*
Woods Hole : Miami	*-0.476*	*0.009*	*38.2*	*18264*	***<0.001***	*-0.442*	*-0.510*
Woods Hole : Varna	*-0.531*	*0.009*	*42.8*	*17932*	***<0.001***	*-0.497*	*-0.565*
Woods Hole : Villefranche	*-0.032*	*0.009*	*2.5*	*19678*	*0.094*	*0.003*	*-0.067*
Miami : Varna	*-0.055*	*0.004*	*8.9*	*28241*	***<0.001***	*-0.038*	*-0.072*
Miami : Villefranche	*0.444*	*0.005*	*64.4*	*27555*	***<0.001***	*0.463*	*0.425*
Varna : Villefranche	*0.499*	*0.005*	*73.7*	*27147*	***<0.001***	*0.518*	*0.481*

Table T-2: Summary of one-sample t-tests for Tajima's *D* for each population.

Population	*t*	*D.o.F.*	*p*	*C.I. (95%)*	
Miami	*-19.4*	*14151*	***<0.001***	*-0.096*	*-0.078*
Sylt	*12.9*	*14166*	***<0.001***	*0.123*	*0.167*
Varna	*-33.1*	*14144*	***<0.001***	*-0.150*	*-0.134*
Villefranche	*68.1*	*14094*	***<0.001***	*0.347*	*0.367*
Woods Hole	*33.4*	*14155*	***<0.001***	*0.366*	*0.412*

Supplement U

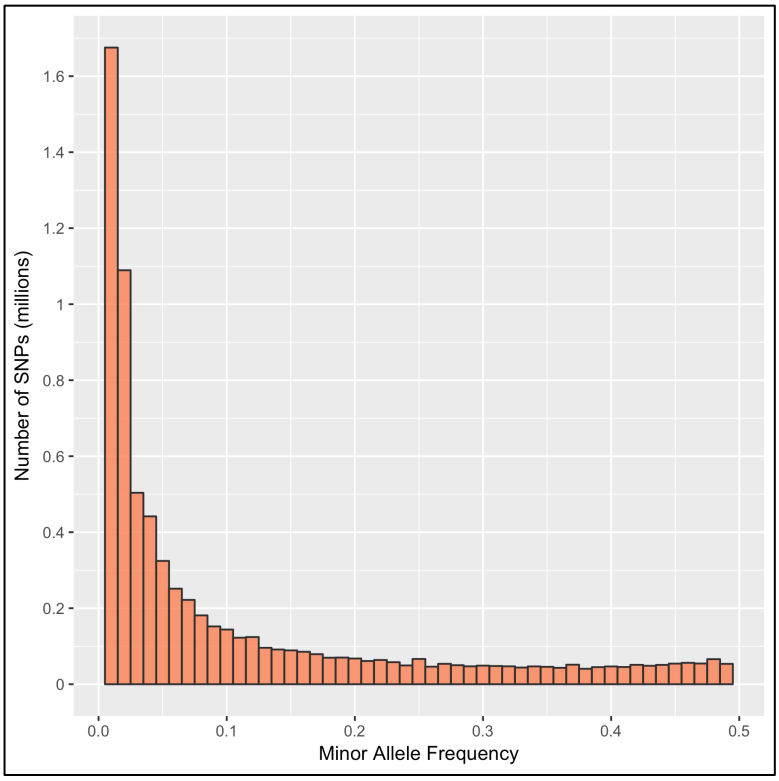

Figure O: Global, folder minor allele frequency spectrum. SNPs with MAF=0.5 are excluded as the minor allele is undefined in this case.

Supplement V

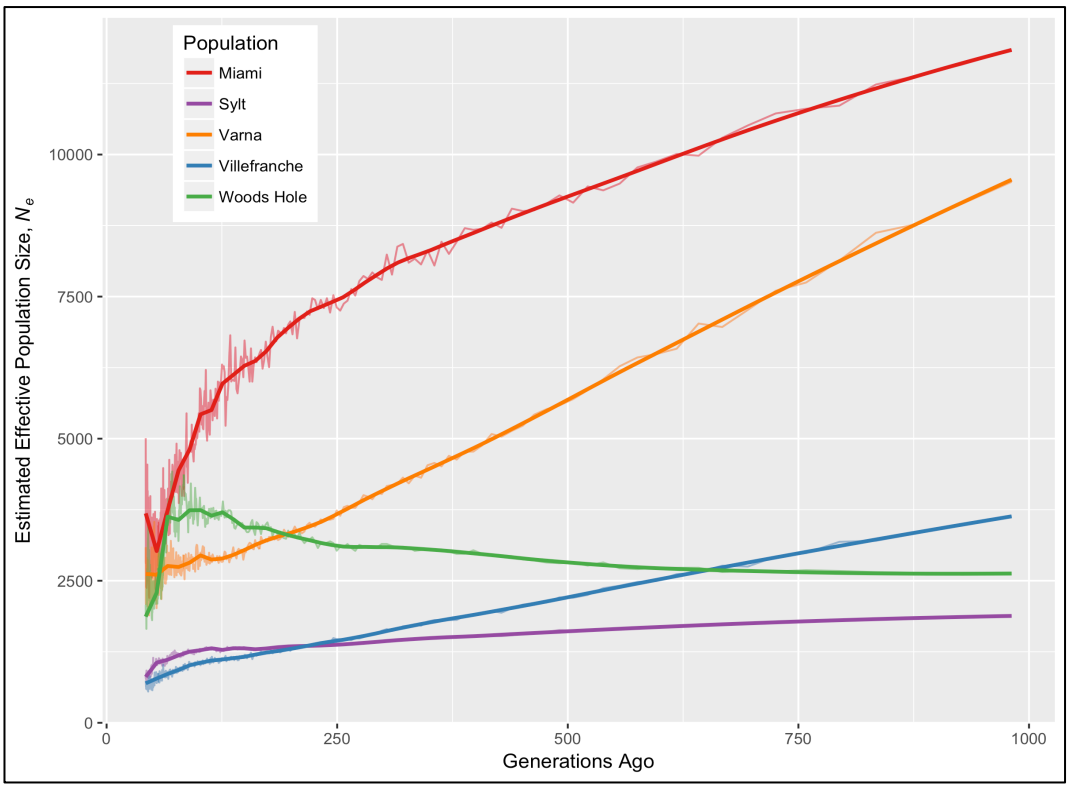

Figure P: Estimated effective population size, N_e, as a function of generations in the past. Estimates were generated from unphased LD data using the method of Corbin *et al.* (2012). Lightly shaded lines link the point estimates of N_e, darker lines represent the running mean of the point estimates.

Declaration of authorship

Hereby I, Moritz Ehrlich, certify that, besides consultation with my supervisors, the Master thesis "*Invasion Genomics: Population structure and diversity patterns in the invasive ctenophore Mnemiopsis leidyi based on whole-genome re-sequencing*" was independently prepared by myself. None other than the indicated resources and references were used. This thesis has never been submitted for a degree or any other qualification at this university or any other institution. The paper format of this thesis is identical with the electronic format submitted alongside it.

Signature Date

Ich erkläre, dass ich meine Masterarbeit „*Invasion Genomics: Population structure and diversity patterns in the invasive ctenophore Mnemiopsis leidyi based on whole-genome re-sequencing.*" selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe und dass ich alle Stellen, die ich wörtlich oder sinngemäß aus Veröffentlichungen entnommen habe, als solche kenntlich gemacht habe. Die Arbeit hat bisher in gleicher oder ähnlicher Form oder auszugsweise noch keiner Prüfungsbehörde vorgelegen. Ich versichere, dass die eingereichte schriftliche Fassung der auf dem beigefügten Medium gespeicherten Fassung entspricht.

Unterschrift Datum